MATE. Multi Aircraft Training Environment

Hauland, G.; Bove, T.; Andersen, Henning Boje; Weber, Steen; Eriksson, R.

Publication date:
2002

Document Version
Publisher's PDF, also known as Version of record

Risø-R-1324(EN)

MATE
Multi Aircraft Training Environment

G. Hauland, T. Bove, H.B. Andersen and S. Weber (Risø)
R. Eriksson and G. Fletcher (DERA)

ESPRIT PROJECT 8982

Deliverable 242
Training Effects Findings and Analysis

AVITRACO	Denmark Project Co-ordinator	
TT&S	United Main Partner	Kingdom
RISØ	Denmark Main Partner	
DASA	Germany Main Partner	
DERA Farnborough	United Associated Partner	Kingdom
DATACEP	France Associated Partner	
DMI	Denmark Associated Partner	
Division	United Associated Partner	Kingdom

Risø National Laboratory, Roskilde
January 2002

Abstract A medium fidelity and low cost training device for pilots, called the Multi Aircraft Training Environment (MATE), is developed to replace other low fidelity *stand-alone* training devices and integrate them into a flexible environment, primarily aimed at training pilots in checklist procedures. The cockpit switches and instruments in MATE are computer-generated graphics. The graphics are back projected onto semi-transparent touch screen panels in a hybrid cockpit mock-up. Thus, the MATE is relatively cheap, it is always available, it is reconfigurable (e.g. between types of aircraft/models to be simulated) and with possibilities for including various forms of intelligent computer assistance. This training concept and the technology are not specific to aviation, but can be used to simulate various types of control panels in different domains. The training effectiveness of pilots' procedure training in the MATE prototype was compared with the effects of traditional training that included the use of real aircraft. The experimental group (EXP) trained the pre-start checklist and the engine start checklist for the Saab 340 commuter aircraft in a MATE prototype. The control group (CTR) trained the same procedures using the aircraft (a/c) for training the pre start and a desktop computer tool (power plant trainer) for training engine starts. Performance on the pre-start checklist was compared in a formal checkout that took place in the a/c. Performance on the engine start procedure was compared in a full flight simulator (FFS). The conclusion was, firstly, that training in the MATE prototype can result in an equally good performance as the existing training (a/c and computer tools), provided that the MATE trainees are given time to familiarise themselves with the a/c. Secondly, training in MATE can result in better performance during dynamic tasks, such as abnormal engine starts. This is promising for the further development of the MATE concept.

ISBN 87-550-3015-7 (Internet)
ISSN 0106-2840

Contents

1 Introduction

The MATE project was initiated to look into the possibility of building a low cost and medium fidelity training device that would be sufficient for training procedures, i.e. one would like to avoid training of procedures in much more costly training environments (i.e. a very high degree of realism). At the same time there is no sufficient low cost training device today that can offer this optimum combination of training effectiveness and cost. Today, much of the initial training is performed in a paper mock-up with photographs of the cockpit, called a paper tiger. Training of more dynamic tasks, i.e. tasks that require system feedback, is performed in various types of desktop computer tools or in relatively expensive high fidelity cockpit systems simulators. Finally the training is finished in real aircraft and full flight simulators. In general terms it appears that the fidelity of training devices for training procedures is either (too) low or (too) high. Extensive use of the a/c for training procedures is vulnerable to costly delays of the training course and ineffective use of training time (e.g. this evaluation was delayed by several weeks because of this). There is therefore a need for a cost-effective training device filling this gap.

A computer generated "virtual" training environment may have the capability of meeting these needs. Such a training device could incorporate many of the *stand-alone* low fidelity training devices, and it could also include parts of the more theoretical computer-based training. Further more, it could log the training activity in various tasks and use this to provide helpful feedback to the trainee. Such a device can have all the aircraft versions (types) needed stored on disk. In sum, a computer generated "virtual" environment would be cost effective in use if it can provide a sufficient quality of the procedural training. When fully developed, such a system could be put to use in all types of training were high fidelity is less important, e.g. flying by instruments, crew resource management (CRM) training and preliminary testing of new instrumentation. From a research point of view, there would be a benefit in the ability to produce many versions of an instrument just by changing computer graphics.

The main issue in relation to effects of training was the system-input device. In a hardware model of a specific aircraft, real aircraft parts are used for switches, controls and instruments. In MATE the trainee had to interact with graphical representations of the real environment. This could have been done through a mouse or perhaps through a VR helmet with glows. However, a high fidelity was required in terms of rehearsing spatial relationships. (The graphical resolution in low cost VR helmets was not high enough at the time to present all the details on the cockpit panels.) A solution with back projection on a semi-transparent touch screen panel was developed. In order to investigate to what extent this was a sufficient solution, it was necessary to determine if two-dimensional picture like representations of switches (and controls) could provide the necessary training for operating real switches (and controls). A literature study (D211) and an experimental pre study (D214, see also D231) were conducted to look into this question. The preliminary answer was that it should be possible to achieve the same effect of training by using touch screen inputs (see Andersen H.B. and Hansen J.P., 1996).

This MATE evaluation had as the main objective to evaluate a complete training environment, consisting of the touch screen cockpit panels mounted in a cockpit mock-up with added paper tiger parts (and demo objects) for the parts of the cockpit equipment that were not simulated using touch screens. To what extent would learning effects of training in the MATE be transferred to the real environment? As a minimum, pilot trainees should be able to manage tasks in the real environment after a relatively short familiarisation time before MATE has proven to be a cost/effective alternative to existing training.

1.1 The Training Environments

The MATE project has collaborated with Skyways, a small regional airline in Sweden (located at Arlanda), and used their training on the Saab 340 aircraft as a MATE demonstrator. The Saab 340 is a commuter aircraft usually set up with approximately 30 seats. There are two versions of this aircraft, the A-type and the B-type. Both types have been used during this evaluation of MATE, but there are only minor differences between these two Saab versions. The training has several parts, but we are exclusively focusing on the *technical training* and the training of engine starts in a *full flight simulator* at SAS Flight Academy in Sweden (also located at Arlanda).

The technical training is aimed at managing the pre start checklist procedures. The technical training taking place at Skyways consists of classroom lectures, various computer based self studies, training in the *paper tiger*, training in the *power plant trainer* and training in the real *aircraft* (see below). The technical training ends with theoretical and practical tests. The practical test is a formal checkout on the pre start checklist in the aircraft (a/c). When this is completed, the training continues in the *full flight simulator* (FFS).

All parts of the checklist procedures studied here are taking place on the ground. Several training environments with various levels of fidelity may be used for this type of training. The following describes the training environments used at Skyways during the MATE evaluation:

- The Paper Tiger: paper pictures/posters with good photos of the real cockpit in life size and with correct spatial relationships.

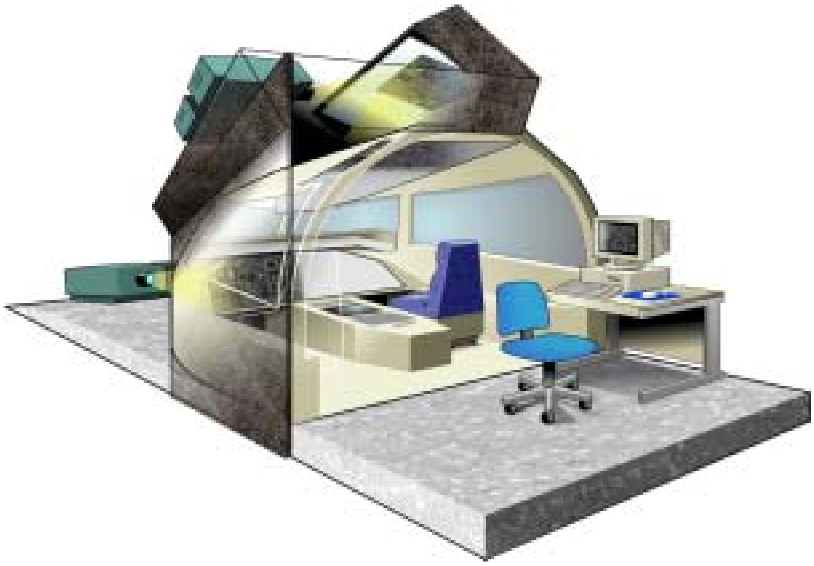
- The Power Plant Trainer: a standard desktop computer and partly a paper tiger. The overhead panel is a paper tiger, but switches related to engine starts are real switches. The instrument panel is not displayed on the computer, instead, single critical instruments are displayed together with other essential a/c information and pedagogical information (e.g. diagram of the fuel flow/propeller pitch).

- The Aircraft (A/C or a/c): the a/c is the aircraft cockpit. This is used for training the pre start checklist only. Trainees are not allowed to start the engines. The aircraft used was an A-type and there have also been some smaller equipment modifications between A-types.

Figure 1: The Saab 340 cockpit in the aircraft (a/c). The cockpit in the full flight simulator (FFS) is identical. Copyright: Saab Service Partner.

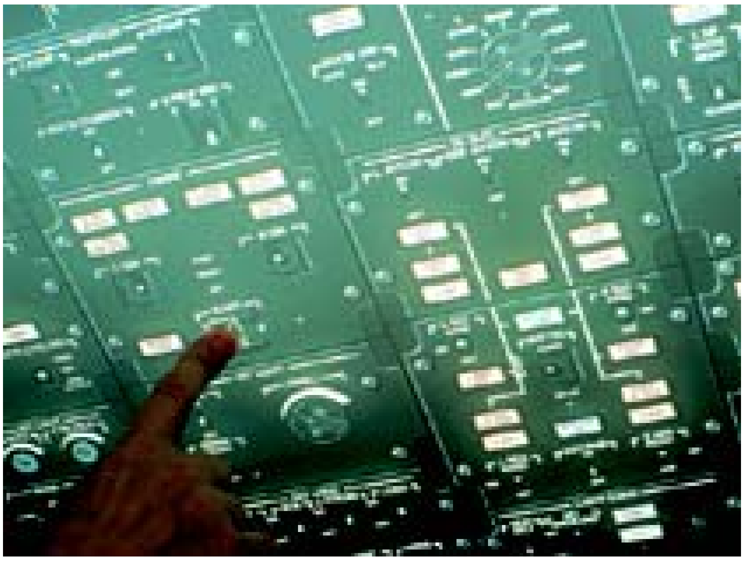
- The Full Flight Simulator (FFS): the physical surroundings are identical to the aircraft cockpit. Trainees now know the pre start checklist (they have been checked out in the a/c), but they perform it as a normal part of the full flight simulation. The sample stops after the engines have been started, i.e. before there is any motion in the FFS. The full flight simulator at SAS Flight Academy is a Saab 340 B-type.

- The Multi Aircraft Training Environment (MATE) Prototype: the size and shape of the physical cockpit mock-up, i.e. the spatial outline of the cockpit walls and roof, is standardised to a hybrid between medium sized commuter aircraft cockpits. The size of the two touch screens is large enough to implement a wide range of cockpit switches and instrumentation. The MATE prototype (located at the SAS Flight Academy in Sweden, Arlanda) was set up as a Saab 340 A-type with real throttles (controls).



<u>Figure 2:</u> the MATE set-up consisted of the cockpit mock-up, the projectors for back projection onto the semi transparent touch screens (overhead panel and instrument panel) and the instructor control station. It also included paper tiger parts (pedestal and side panels) and a demo object (oxygen mask).

<u>Figure 3:</u> the MATE cockpit. Touch screens with back projected graphics: the overhead panel and the instrument panel. Real throttles were used in this prototype. Visuals are not a part of MATE because they are not required for procedural training.



<u>Figure 4:</u> *detail from the touch screen with back projected computer graphics of the overhead panel. All switches (including spring loaded switches and switches with guards) and selectors can be moved (sometimes with a bit of difficulty) by using "one finger touch" instead of a grip (like in the real aircraft). Sounds (e.g. a click from moving a switch) and lamps, including various aural and visual cautions and alarms, are simulated together with the graphics.*

1.2 The Checklist Samples

The checklist, including the pre start and engine start checklists, are divided into one normal and several abnormal procedures for operating the a/c, i.e. procedures for normal and for a range of abnormal conditions. The abnormal checklist describes what the pilot must do if certain symptoms of a/c malfunctions can be observed, e.g. "the hung start" (the engine gas turbine has stopped accelerating) or "the hot start" (an over-temperature in the engine between the gas and the power turbine). The normal checklist consists of items (main headings) with keywords to be performed during normal operations. The extended checklist describes in detail each point on the normal checklist (several sub headings). These descriptions must be learned by heart. The normal checklist, i.e. the short version with main headings, is the one that pilots use in line operation. The trainees had both checklists available.

Instructors may sometimes have more *practical* ways of applying the normal checklist to line operations than actually described in the extended checklist. At the same time, trainees are required to follow the extended checklist. Thus, the trainees may sometimes experience conflicting opinions on how the checklist should be applied. Nevertheless, during this evaluation the written official Skyways extended checklist (Skyways version 960901, see appendix 11.1) was regarded as a description of the only correct procedural performance in the cockpit. This checklist is undergoing a constant review, but it is an official company document until it is officially changed. There is one instance however, where there was reason to believe that differences between instructor practices and the extended checklist have had an impact on the trainee's opportunities to learn a specific flow pattern (the *avionics/inverters* sequence in the pre start checklist). Errors related to this specific sequence are therefor disregarded. There are also other variations in the correct checklist procedure relevant to our samples. First, there is a difference between the Saab 340 A- and B-type. This has an implication on one point within our pre start sample (the *auto coarsen* system). Secondly, there is a difference in the pre start checklist dependent on the type of power source in use (battery start and bleed leak test, the latter is not performed). The trainees are briefed in detail about these variations and these differences should therefore not have any impact on the learning of the correct checklist procedures.

1.2.1 Pre Start

The purpose of the pre start checklist is to prepare for the engine start and take off. Performance during this sample is individual and performed by the first officer (Right pilot / R). It is not required that the captain is present (Left pilot / L) and call outs are therefore not required. Both trainees in the crew were present in the a/c during training of the pre start checklist, but the checkout was performed individually with the instructor in the L seat, evaluating one trainee at the time.

The task during the pre start checklist is *static* in nature; i.e. it is simply checking the cockpit configuration in relation to the (extended) checklist. The cockpit configuration is already close to the configuration described in the pre start checklist (the way the crew left it after the last flight of the day), so the number of configuration changes that the trainee must make is relatively low. Thus, in order to observe any errors, it was necessary to try to increase the potential for errors. This was done by changing the default configuration before the trainee entered the cockpit (the inserted error potential/pre start, appendix 11.2).

In this evaluation we used two samples from the pre start checklist: (a) the *training sample* and (b) the *checkout sample*:

a) <u>Training sample</u>: the range which this sample covers is touch screen specific, i.e. we only score performance in relation to the functionality represented on the MATE touch screen version of the overhead panel (items in appendix 11.1/the pre start part). The training sample is a part of the checklist that is (almost) identical between aircraft version A and B. In this sample, we do not collect data on performance in relation to items that are checked visually, because checking a specific item is not accompanied by actions that we can observe (e.g. eye movements). The training sample is recorded in pass 2 – 4.

b) <u>Checkout sample</u>: this covers the complete pre start checklist. It includes the training sample and it has additional checklist items before and after the training sample, i.e. outside the camera angles and not relevant to the touch screen part of the MATE checklist (i.e. the operational definition of the Skyways checklist, appendix 11.1). The additional items (not included in the appendix) are related to (physically) checking the emergency and safety equipment (the *minimum equipment list / MEL*, the *oxygen system*). After the checkout, the instructor debriefs the trainee and makes notes (appendix 11.5) about what the trainee checked. This includes asking about what lamps were out of order and to what extent MEL items (the torch) were available and working. The checkout sample is recorded in pass 5 – 7.

1.2.2 Engine Start

The operational definition of an "engine start trial" (normal and abnormal) was: the attempt of one crew combination to start either the left or the right engine. Several start attempts (usually not more than two attempts) on one engine is part of the <u>same</u> engine start, provided that the failure to start the engine is due to the crew performance deviating from the checklist. If the failure to start an engine is due to the error potential inserted by the instructor (abnormal engine starts; hung or hot start), or to the fact that the instructor commands a new start, then the second attempt to start that same engine is considered a <u>new</u> engine start. If there is a shutdown for reasons other than the inserted error potential or instructor orders (i.e. it is not a new start), then the performance in the repetition of checklist items up to the point where this start attempt failed is disregarded. A new engine start however, should always be performed (and therefore scored) from engine checklist item no. 1. (See the analysis chapter for details).

The following describes the sample from the checklist related to engine starts only:

Engine Start Sample

From *Start of Engines* (engine checklist item no. 1, appendix 11.1/engine start sample) to *After Engine Start*. The crew performs this sample together. It takes two persons to start the engines in accordance with the checklist, i.e. the motoring start. The task is dynamic in that the crewmembers must co-operate, interpret various instruments and take actions on the basis of this information. Thus, an understanding of a/c subsystems and how they interact is necessary. Any malfunctions must be diagnosed and acted upon within seconds to avoid damage to the engines.

Motoring

Motoring is the name for ventilating (motoring) hot gasses/fuel away from the engine by engaging the starter without supplying fuel or ignition. Motoring is done as part of a normal start procedure to be sure that there are no fuel gasses or high temperatures in the engines before the engines are started. Motoring is also done if the temperature between the two engine turbines approaches a critical value. These expensive turbines can be seriously damaged if the temperature (inter stage turbine temperature, ITT) is too high (960 degrees C or more). Motoring is performed by holding the spring-loaded starter switch in the L (left) or R (right) position, dependent on what engine the pilot will start. The condition lever must first be in fuel off and the relevant ignition switch must be off. (The ignition switch is guarded, so the guard must first be opened.) This will not only disable ignition, it will also disconnect the hold circuit that otherwise would continue to provide power to the starter. Thus, the pilot has direct control with the starter: it is engaged as long as he is holding the starter switch. It is therefore necessary to time the motoring sequence. During a normal start, the conditions described for engine temperature can be reached within a relatively short period of time. Thus, to save the starter and to save time (i.e. a new starter sequence can not be initiated within the next 3 minutes), the engines are started directly from motoring on the remaining part of the starter sequence (max 70 seconds): the captain orders “fuel on” and the first officer sets the condition lever to the start position and calls out “fuel is on”. Immediately after this (no later than 2 seconds), the captain enables the ignition and places his right hand on the condition lever (to be prepared for a shutdown if there are any abnormal signs, e.g. overheating between the turbines).

2 Method

2.1 Hypotheses

Since the pre start checklist represents tasks that require very little feedback from the aircraft system, we do not expect that the experimental group should perform better than the control group, but at least that the experimental group is able to perform the checklist within the required limits. The Engine Start tasks require interpretation of system feedback and it is expected that trainees from the experimental group will have a benefit from training in an integrated and interactive environment, thus perhaps perform better than the control group on these tasks. The experimental group should at least perform as well as the control group.

The main hypothesis was that (a) the MATE trainees (the EXP group) would only initially perform the pre start checkout worse than the CTR group, i.e. in terms of time and/or errors, when transferred from MATE to the real environment and (b) that MATE trainees would perform engine starts better than the CTR group when the groups transferred to the real environment.

2.2 Subjects and Instructors

The subjects were 20 trainee pilots recruited by Skyways. Two subject groups, with 10 trainees in each group, were undergoing the Skyways training course on the Saab 340. It was not possible to influence how the groups were put together, since the courses at Skyways were set up before the planning of the evaluation. The first group (October-December, 1997) was the control group (CTR), i.e. conventional training. The second group (February-July, 1998) was the experimental group (EXP), i.e. MATE trainees (the MATE prototype was operational just in time for the second course). All trainees were promised full anonymity. All AV recordings will be treated with confidentiality and subsequently be erased.

Background data for the control (CTR) and experimental (EXP) group				
	Age (years)	Paper tiger (hours)	Power plant* (hours)	Mulit-eng. (hours)
CTR	31.2 (27-37)	6.6 (4.5-11)	3.8 (3-5)	499 (50-1100)
EXP	35.5 (29-52)	8.6 (3-13)	3.6 (2-8)	525 (0-1300)

*Normal engine starts only

Table 1: most trainees were in their early thirties, they had the same possibility for free play in the paper tiger (training device), and approximately the same amount of training in the power plant trainer (desktop tool for training engine starts). Multi engine aircraft may be an indicator of previous experience with carrying out checklists (but it is not an indicator of total flight hours).

There were 7 instructors involved in this evaluation: two instructors during a/c training and the checkout, one instructor during MATE training and four instructors allocated to FFS turnus instruction (the FFS was also used during the night). The instructors had different experience with Saab 340 courses: one a/c instructor was the chief instructor (responsible for the whole course) and the other a/c instructor was a captain trainee. The MATE instructor had just been checked out of technical training himself, and he was actually a trainee in the CTR group. One of the FFS instructors was also new in that position.

2.3 Quasi Experimental Design

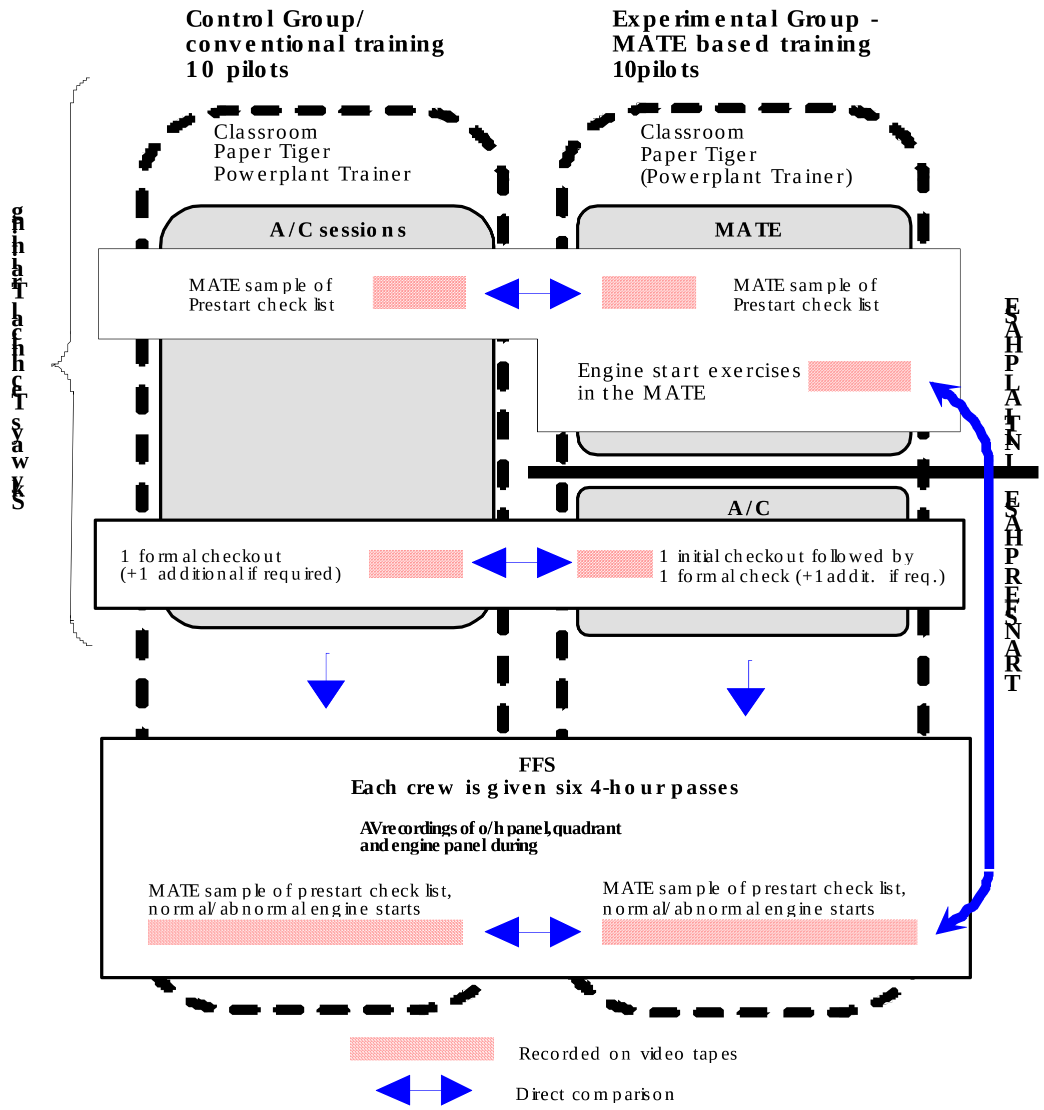


Figure 5: the quasi-experimental design. The main comparisons between the control group (n=10) and the experimental group (n=10) were made in the a/c checkout and in the first three passes of the FFS training.

Each *pass* contained all the 10 trainees, i.e. 5 crews in two combinations: both trainees should try both seat positions. One pass contained 10 *sessions*, each session representing one crew combination. Each trainee wrote down his unique *crew letter* on the back of his checklist, holding his letter in front of the video camera just before the session started. The instructor (was supposed to) state the *pass and crew letter* into the microphone, and also to write this (plus the date) on pre printed labels for the various videotapes and onto the score sheets (instructors were briefed, see also the what-to-do-list, appendix 11.4). During pre start, which is mainly done on an individual basis, this detailed labelling system (i.e. crew letters) was not strictly necessary, as it was always the R trainee that performed this checklist. During engine starts however, the trainees switch positions between sessions. The performance of individual trainees (as opposed to the crew) was not a major interest, but it was often difficult to know how far a particular trainee had come in the training (e.g. one trainee could be sick, i.e. trainees in a crew were not necessarily in the same crew throughout the course. Thus, this labelling system was of great help when analysing the data). One crew combination in one session could start an engine 2 times if it was a normal start (both L and R engine) or 3 times if it was an abnormal start (one extra attempt on one of the engines). These engine starts were labelled *trials* (see the definition of an engine start. It could be maximum 3 trials within 1 session). Thus, trial-labels were only used during engine starts.

A pass number represented the n^{th} time of training the checklist sample. During technical training, the CTR group had 5 passes in the a/c. The first pass was a familiarisation pass, pass 2-4 were the training passes, pass 5 was the checkout pass. The EXP group had the same number of passes in MATE, but with checkout pass number 5 in the a/c. The EXP group was guaranteed more checkout passes if needed, but the trainees were not told about this guarantee before their first checkout attempt. The EXP group needed two more a/c checkout passes, i.e. pass 6 and 7 (pass 7 with only two trainees). All trainees in the EXP group had one additional familiarisation pass in the a/c (pass 8). In the FFS training, which combined the pre start and engine start samples, the CTR group had 6 passes and the EXP group had 7 passes. This last pass was due to new regulations (flying in bad weather/low visibility). Pre start was not emphasised by the instructors during FFS passes (at this point in time, all trainees were checked out from the technical training). Engine starts were emphasised in the 3 first FFS passes, but became less important at the end of the training (where flying the a/c was the main issue). The insertion of abnormalities during engine starts was done in pass 2 and 3 (dependent on training progress, as evaluated by the instructors). The MATE evaluation has focused on these first three passes to minimise training effects and to focus entirely on the pre start and engine start samples. Thus, there were three relevant passes in the full flight simulator (FFS), pass 1-3.

1.1.1.1 PASS

Figure 6: There was a total of 7 passes analysed for both the CTR group and the EXP group (and 3 extra passes for the EXP group). One pass contained 10 sessions. Each session represented one crew combination performing one of the checklist samples once (pre start or engine start). A crew combination was identified from a two-letter code: the left letter position (i.e. role of the captain) represented the left seat and the right letter position (i.e. role of the first officer) represented the right seat (seen from the jump seat back in the cockpit). Trainees switched places between sessions. The control group had letters A-J, the EXP group had letters K-T. Each engine start session contained 2 trials if it was a normal start and 3 trials if it was an abnormal start. (Pre-start sessions did not contain trials.) Trials were labelled L (left) or R (right), start number 1 and start number 2.

The CTR group followed a traditional Skyways Saab 340 course, consisting of classroom lectures and training in paper tiger and power plant trainer (followed by FFS training). The CTR group had four training passes (familiarisation and pass 2-4) in the a/c and one checkout pass (pass 5). The EXP group had the course a couple of months later, with the same instructors (plus one new MATE instructor and one new FFS instructor) and at the same location at Arlanda in Sweden. The difference was that (1) the training in the a/c was replaced with the training in the MATE prototype and (2) the training of abnormal engine starts in the power plant trainer was replaced with training in the MATE prototype. Due to specific pedagogical tools (fuel diagrams) implemented in the power plant trainer, the EXP group had limited training of normal engine starts in the power plant trainer instead of in the MATE. The training of normal engine starts was performed with supervision from the instructor, i.e. the instructor demonstrated the power plant trainer. Both groups should have four passes with engine start training, i.e. two normal and two abnormal passes. The EXP group had two passes with abnormal engine starts in MATE (in addition to the 4 pre start passes). The amount of training (in the a/c, the power plant trainer or the MATE) was equal for both groups: a training pass in the a/c or in the MATE was fixed to 20 minutes. An engine start pass in the power plant trainer or in the MATE was fixed to 60 minutes. Both groups were free to use the paper tiger as much as they wanted. (They had access to this device also during the evenings.)

Performance on the pre start checklist was compared in the a/c without any cockpit familiarisation. This was the first time the EXP group saw the real Saab 340 cockpit. The comparison of engine starts (normal and abnormal) was done in the FFS. The performance was video recorded and the instructors were asked to complete the instructor score sheets (appendix 11.6). During the checkout, instructors also completed instructor notes, i.e. a specification of the types of errors and the time to complete the pre start checklist (appendix 11.5). When the checkout was completed, the trainees were said to have officially completed their technical training. This did not mean that the pre start checklist was disregarded during the next part of their training, but it was a requirement to have passed the technical training before being allowed to operate the FFS. We continued to A/V record the pre start sample in the FFS. For the engine start sample (immediately after the pre start sample), we recorded an additional tape of the instrument panel and the area around the throttles (referred to with various degrees of accuracy: controls / quadrant / throttles / pedestal). Thus, we had one video recording (with audio) of the overhead panel and one video recording (with audio) of the instrument panel/pedestal in the FFS. In addition to this, both instructors and MATE trainees were asked questions (questionnaires, appendix 11.8) concerning the training in MATE (and background data).

2.4 Procedures

PASS		**DESCRIPTION**	**ENVIRONMENT**
Pre Start	Pass 1	Familiarisation *(not analysed)*	A/C or MATE
	Passes 2 - 4	Training (training samples)	A/C or MATE
	Passes 5 - 7	Checkout (checkout samples)	A/C *(comparison)*
Engine start	_Passes 2 - 4*_	Training normal and abnormal starts *(not analysed)*	Power Plant Trainer or MATE
Pre start & Engine start	Passes 1 - 3	Combined passes. (Simulator training and Skyways rating.)	FFS *(comparison)*

<u>*Table 2:*</u> the aircraft (a/c) were accessed at Skyways' terminal at Arlanda airport in Sweden (the parking place or the hangar). The Multi Aircraft Training Environment (MATE) and the Saab 340 B full flight simulator (FFS) were accessed at the SAS Flight Academy, also located at Arlanda.

**No recordings were made from the training in the power plant trainer. The MATE trainees had two passes (60 minutes each) with abnormal starts in the MATE, but these passes were not analysed due to resource problems and low priority. The main comparison was done with a/c checkout and FFS data.*

The first passes in both the CTR group and the EXP group were familiarisation passes that were not analysed. The pass number refers to the nth time of training (MATE, a/c or FFS), i.e. there are 10 trainees forming 5 crews (crew combinations) performing 10 sessions (various number of trials) in each pass (in the FFS, Skyways refer to passes as sessions). The demands to speed and accuracy increases pass by pass, and so does the inserted error potential. Pass 1 is a familiarisation pass for all personnel (including testing data recording) and trainees.

Pre start
Pass 1 was a familiarisation pass for all involved personnel (trainees, instructors, researcher, and equipment). Passes 2 – 4 were the training passes. Pass 5 was the first official checkout pass, passes 6-7 were two more checkout passes for the EXP group (if needed).

Engine start
Engines are not started for training purposes in the a/c. The FFS is used for this part of the training. There were 6 passes in the FFS for the CTR group. Due to new regulations, there

were 7 passes in the FFS for the EXP group. If the instructor thought that a trainee had reached an adequate level of performance on certain checklist items, these items were no longer stressed. Thus the relevance of the FFS training to this test decreased towards the end of the FFS passes. Also, the FFS passes were both training and tests, i.e. the instructor would help the trainee, but he would also set a mark on his performance.

To maintain focus on the pre start and engine start, and to reduce learning effects from training in the FFS, we will focus on the three first FFS passes. These three FFS passes contain the pre start checklist. This will be recorded as a possible way of controlling how well instructors are able to predict performance of the trainees. The interesting part in the FFS is the engine start, following directly after the pre start. The engines are started as a crew and each individual crewmember must try both roles in the crew (i.e. the captain and the first officer). Trainees change seat positions between sessions. However, due to the expensive simulator time and variations in performance, this set-up may be changed unsystematically by the FFS instructor. The most interesting figures will therefore be the total number of errors (based on performance deviations from the checklist) observed in each pass.

2.5 Apparatus

Skyways use real aircraft for training of the pre start checklist. Studying performance of trainees in the real cockpit turned out to be a practical challenge. Whenever a Saab 340 is temporarily parked at Arlanda (for maintenance and preparations between flights), it is used for training. Sometimes training in the cockpit and maintenance were in conflict, thus resulting in a delay (or cancellation) of training. This means that instructors and trainees have to be stand by at the airport (or close) for longer periods of time. Thus, the researcher must have mobile equipment that can be mounted without any permanent arrangements and adjusted very fast. We have used a portable monitor with a wire extension. This is important, because the researcher can then monitor recording from the cabin instead of the cockpit. We have used two 12V power packs (batteries), a standard video camera with batteries. (Several backup batteries). We had to have a lot of backup power. There is no DC 220V available in the ground power unit (and we did not know if the a/c was parked at the gate/parking place or in the hangar).

One SVHS-C camera on a removable bracket was used to video and audio record the performance on the overhead panel in the a/c and in the MATE. Due to practical limitations the camera had to be placed on the floor, close to the pedestal, and record the panel in a landscape format (i.e. tilted). One digital camera (DVD) was used to record performance on the overhead panel in the FFS. The digital camera was small enough to be mounted on the wall between the R pilot and the instructor seat. The small LCD monitor on the camera could be adjusted so that the instructor was able to monitor the recording. The perspective was the same as used in the a/c and in the MATE. The same type of camera was used in the MATE, mounted in the roof of the cockpit mock-up, to record the pedestal (and parts of the instrument panel). An ordinary VHS black and white camera, mounted in the roof above the instructor, was used to record the pedestal in the FFS.

Video and Audio Analyses Techniques

The various video formats (DVD, SVHS-C, SVHS and VHS) were copied for backup onto ordinary SVHS and VHS tapes. Copies of the backup tapes were also made and sent to a partner in the MATE project, the Defence and Evaluation Research Agency (DERA) in the UK, for independent descriptive scoring of the same data set. A descriptive score sheet was first developed in collaboration with DERA. The score sheets contained the operational definitions of the Skyways extended checklist (appendix 11.1).

The objective pre start data consisted of one video recording of performance on the overhead panel per trainee. The objective engine start data consisted of two simultaneous video recordings of the performance, one for the overhead panel and one for the instrument panel / pedestal, per crew combination. Synchronisation of engine start tapes was not necessary, since actions could easily be followed between videos and the researcher could only attend to one videotape at the time. Thus each tape was viewed separately for the purpose of scoring, although the descriptions were merged into one descriptive score sheet. One descriptive score sheet was completed for each trainee in the pre start part (a/c and MATE) and one for each crew combination in the pre start + engine start part (FFS). A SVHS / VHS video machine with shuttle control was used for playback of the backup tapes to be analysed. Describing the raw data from video was done close to real playback time (for each video). Two monitors, one 28’’ (including digital still-picture, some details had to be enlarged) and one 14’’ were used. Descriptions were made regarding every deviation from the operational definition of the extended checklist (appendix 11.1).

Sound was recorded on all videotapes, i.e. two sound recordings from the FFS. Due to the location of microphones, the sound quality varied between the recordings, but an adequate audio playback was produced. However, the sound was sometimes heavily disturbed by a/c systems (e.g. the a/c hydraulic system, interference from electrical systems in the FFS disturbed some of the EXP recordings). An equaliser was therefore used to filter specific frequencies. Both the complete set of checklists (normal, extended and abnormal) and a detailed enlarged picture of the overhead panel were available to the analysts at all times. The picture was often compared with the video recording, since lighting conditions on recordings varied. It was sometimes difficult to decide about the position of a specific switch by looking at the video picture only. Attention was given to audio and video recordings simultaneously, to actually hear the *click* when a switch was moved (the MATE also simulated sound).

3 Analysis

The MATE (the MATE trainer prototype was the independent variable) was evaluated against the *conventional* training environments: the *a/c* and the *power plant trainer*:

Three analysts have been involved in the data analyses, two from Risø and one from DERA. All three analysts spent time (approximately 5 days) learning the checklist and understanding some basic features of the a/c system interactions. There was complete agreement between DERA and Risø on the raw descriptive scoring of data from video, i.e. the actions deviating from the actions described in the operational definition of the extended checklist. The categorisation of raw descriptions into errors was discussed between DERA (UK, Farnborough), Risø (DK), and Saab 340 instructors at Skyways and at the SAS Flight Academy (SE). There has not been any independent error categorisation and therefore no test of inter-scorer reliability (the project was delayed and the categorisation system was not completed).

A distinction was made between *objective* data and *subjective* data: the objective data were based on recordings of subjects’ behaviour and subject/system interaction, i.e. the raw data can be considered to be relatively free of judgements. The objective performance data from all three conditions (a/c, MATE, FFS) were first described in relation to the extended checklist. These descriptions were categorised into errors so that the two groups could be compared statistically. The subjective data were the instructors’ ratings of trainees’ performance and the instructors’ and trainees’ opinions of training in the MATE.

Overview of Data:

- Audio/Video recordings (training sample passes 2-4, checkout sample passes 5-7) and instructor notes (pass 5) from the pre start checkout and the checkout debriefs: errors related to checklist *actions/checking* various items, *callouts*, *time* to complete the checkout and *role* in the crew.

- Skyways' evaluation of the completed pre start checkout (approved/not approved).

- Audio-tape from the additional familiarisation pass (pass number 8).

- Instructor score sheets (their subjective impressions of both CTR and EXP).

- Instructor opinions of MATE as a training device

- Trainees' opinions of training in MATE.

Objective and Subjective Data:

When the raw data is in a form that can be said to be free of interpretation, we refer to it as *objective* (e.g. from the A/V recordings). When the data represents an impression/opinion of some sort, we refer to it as *subjective* (e.g. Instructor score sheets). The objective data is categorised into various types of erroneous performance in terms of time and errors. The following describes the analysis from the *description* of raw data to the *error categorisation*:

3.1 Description of Data

Observable crew performance deviating from the performance described in the Skyways Saab 340A extended checklist (version 960901, appendix 11.1) were described *action by action*. These descriptions of deviations were later categorised into errors. The Saab 340 instructors were consulted when deciding what type of deviations should be considered erroneous.

Observable System Variables

There were no system logs available in any of the training environments (no log facilities were implemented in the MATE prototype). The relevant system variables were therefore logged by means of A/V recordings (appendix 11.1/Engine Start Sample) of instruments, lamps and alarms (aural cautions and warnings). In addition, the instructor wrote down task specific variables (the inserted error potential and the type of starts).

Observable Crew Variables

During the pre start sample, individual assessments were made of switch actions on the overhead panel and the total time to complete the pre start checkout. The remaining variables apply to the engine start part of the checklist. The crew (i.e. two trainees) performed the engine start checklist.

Actions on Checklist Items:
All the actions that deviated from the extended checklist were described. The descriptions will be kept as a long-term record (the videos will be erased). Chunks of actions in the extended

checklist were broken down into single observable actions and related to each observable change in the configuration of a panel. Actions that deviated from the checklist descriptions within a main checklist heading were described in relation to the end result of those actions, i.e. position changes, rather than other aspects of how these actions were carried out, or in what sequence. However, mismatches between the checklist sequence of main headings and the observed sequence of main headings were described.

Actions fall into two main categories: non-verbal actions and verbal actions. Non verbal actions were (hand / finger) movements related to the cockpit configuration of various switches, including guarded and safe tied switches and handles (for securing switch and handle positions), selectors (e.g. the voltage selector) and controls (i.e. the throttles). It also included finding and reading an abnormal checklist, and giving the disconnect sign. Some actions did not leave any observable result in the interface (i.e. change of positions): spring loaded switches, like the test switches and the engine start switch, will always have a centre position. The result of actions on these switches was observed in relation to direct system consequences of these actions. Mandatory verbal actions were the uttered callouts and commands. The callouts, i.e. the standardised verbal statements, are mandatory formal ways of communicating about the a/c system status at critical points during the execution of the checklist. A callout is usually a request or a positive feedback on a request. There were no formal callouts at Skyways that trainees were obliged to use if the system premises for a callout were untrue. *Commands* (i.e. direct orders) were included in the *callout* term (because it was only one actual *command* in the engine start sample).

Finally, there were many actions related to visual observations to ensure that the checklist items were in order. Visual activity was not measured, but the instructor made notes during the a/c pre start checkout concerning the checking of lamps (malfunctioning of light bulbs). The instructor also made notes concerning the checking of the minimum equipment list, MEL (e.g. that the torch was there and that it was working). The notes were confirmed by asking the trainee about the system status at various points in time during the checkout. This took place during the debrief in the cockpit, immediately after the checkout.

Time to Complete the Crew Actions:
Time was (a) a separate measure of performance during the complete pre start checkout (in relation to the official Skyways minimum requirement of 15 minutes). Time was also (b) measured to evaluate the correctness of specific actions (complying with time limits) during engine starts. The latter (b) will be referred to as *timing*.

Role of the Crewmember
The seat position determines the role as a crewmember: the L (left) seat represents the role of the captain and the R (right) seat represents the role of the first officer. The checklist determines which crewmember should perform what actions. The conformity to these requirements was observed. The seat position of the trainee performing the action was scored if their role in the crew was reversed, i.e. R pilot performing checklist items that the L pilot should have performed and the other way around.

3.2 Categorisation of Erroneous Performance

There were two types of objective performance measures: (1) time to complete the pre start checklist during the checkout and (2) number of checklist errors (type and mode).

<u>Time for completing the pre start checkout (1):</u>

Using more than 15 minutes to complete the pre start checklist does not have any direct a/c system consequences, but could potentially cause inconvenience because the scheduled departure could perhaps be delayed if the crew spent too much time on completing the pre start checklist. The definition of a 15 minutes limit is a company requirement that may vary between airlines (the limit for passing the same technical training in SAS is 20 minutes). Exceeding these 15 minutes can therefore not be considered a checklist error. This measure of time is a separate indication of the quality of the performance in relation to Skyways standards. As such, this time limit is less important than the checklist errors.

<u>Checklist Errors (2):</u>

The operational definition of the checklist describes verbal actions (related to callouts, including commands) and non-verbal actions (related to the cockpit configuration of switches etc.) in relation to: a/c system status, time limits and the role of the crewmember.

The concept of *error* was related to the operational definition of the checklist, the observed a/c system status, the observed timing, the observed role and the inserted error potential (and instructor commands). The inserted error potential (see appendix 11.2) describes the a/c system variables that have been manipulated during the evaluation, i.e. it constitutes a potential for errors. This potential was pre defined, i.e. the specific cockpit configuration deviating from the default was decided before the evaluation (it was an attempt to make the various training tasks more difficult, but the instructors did not always comply with this set-up). Actions deviating from the correct application of the checklist will be erroneous and each observable deviation will count as one error. If some consequential actions can be regarded as subjectively correct due to a misinterpretation of the system status, they are still objectively wrong, e.g. handling an erroneous shutdown correctly is still erroneous.

Data was not recorded for interpreting any cognitive aspects of errors, e.g. discriminating between perception (e.g. seen/not seen) and higher cognitive processes (e.g. intention/understanding/knowledge/diagnosis). Observations were made exclusively of the results of behaviour, i.e. if the observed actions deviating from the correct cockpit configuration, callouts or the role of the crewmember. Also, the categorisation of errors was not related to specific types of switches. (Only the descriptive scoring was related to specific single checklist actions, appendix 11.1.) Each erroneous action was counted as one error. Potential consequences, i.e. new checklist actions and error potentials, of trainee induced errors were disregarded. However, sometimes it was difficult to know if the start was objectively defined (i.e. not trainee induced) as normal or abnormal, due to the incomplete instructor score sheets. When this was the case, assumptions have been made based on conversations between instructors and trainees to determine if it was an abnormal start (objectively defined) or if the trainees just thought so.

The categorisation system used for classifying and counting observed behaviour as erroneous (see e.g. Rasmussen, J., 1982, Reason, 1990) had two dimensions: specific *error types* and general *error modes* (detailed explanation below). The words *type* and *mode* are both used as names for an error category, but these two dimensions represent different levels of abstraction. The main rule for counting was one error type and one error mode per action deviating from the operational definition of the checklist (score sheet, appendix 11.1).

The three error types were (a) cockpit configuration errors (config. errors), i.e. the position of switches and controls in relation to the checklist and the system status, (b) callout errors, i.e. erroneous phraseology or a mismatch with the system status and (c) role errors, i.e. related to whom was performing the action. These error types can also be described in functional terms, i.e. the three general error modes: (a) sequence of actions (flow), (b) missing actions (omissions) and (c) wrong actions (commission). The following section will first describe error types and then describe error modes. (See table below.)

1. The three types of errors: domain specific terminology was used to name the specific error types (configuration and callout). The *role* category (grey area) is an additional category not related to the immediate consequences on the a/c system status.

2. The three general error modes: these were used for describing the formal characteristics of the error types. The *flow* error mode has *timing* as a sub category. The *omission* error mode has *transition* as a sub category. Note that *role* (grey area) is not classified in any of the modes. Also note that one erroneous callout can be counted in two error modes (two grey cells).

	ERROR TYPES		
ERROR MODES	**Config. errors** Switches/controls in relation to the checklist and the system status.	**Callout errors** Orders/callouts phraseology in relation to the checklist and the system status.	**Role errors** Reversal of whom is performing the checklist actions.
Flow			
Flow (timing)			
Omission			
Omission (transition)			
Commission			
Total			

Grey means AND, i.e. additional error.

Table 3: Categorisation of performance from the score sheets (i.e. the operational definition of the checklist, appendix 11.1). The error categorisation system has two dimensions: (1) three domain specific error types and (2) three general error modes. (Appendix 11.3)

The following section will elaborate on the error types and modes.

Error types

Non verbal actions were referred to as (domain specific) *configuration errors*. Configuration errors are any results of actions on cockpit switches and throttles/controls deviating from the actions (results) described in the extended checklist.

Verbal actions were referred to as (domain specific) *callout errors*. One question pertinent to the scoring of callout errors was whether or not a callout should be interpreted as correct if the meaning reference was judged as *clear enough*, or should the strict form of the checklist callout be used as the only acceptable callout, i.e. without any interpretation? The latter option was regarded as the most adequate for scoring and also preferable because one of the main purposes of the checklist is to formalise communication (to increase safety/reliability). However, problems with the sound made it difficult to discriminate between smaller phonetic details (due to hydraulic noise and electrical interference) and language problems (trainees' command of English/translation problems for the researchers; cockpit communication in Swedish).

Callouts were not analysed on a higher resolution than the level of propositions. Thus, each callout statement was one unit of analysis. If a *callout* came in the wrong place within the sample, it was one *flow error*. If it did not come at all within the sample, it was an *error of omission*. A common deviation was that something was missing from the complete callout as defined in the checklist. Seen in isolation, this may constitute an *error of omission*. However, here the proposition (call out) is seen as one action and it is wrong in relation to the a/c system, hence an error of *commission*.

The callout meaning reference can constitute an error of commission in two different ways:

1) The phraseology can be wrong (semantics, syntax), i.e. it is deviating from the checklist phraseology. The meaning reference can be misunderstood because the callout phraseology has changed.

2) The meaning reference of the phrase can be wrong because the system status does not match this reference; i.e. the premises for the callout have not been met. Thus the correct performance would be the absence of a call out (i.e. *checked* means that something is OK) and start investigating the a/c system malfunction.

Distinctions were not made between these two aspects of incorrect meaning references when counting callout errors. There was one *callout error of commission* if at least one of the following criteria were met: (1) if the analyst found any deviation from the exact checklist phrase. Practically, this was a scoring of morphemes (meaningful part / the stem of a word), and (2) if the system conditions that the callout was referring to were untrue. Practically, this scoring was limited by our ability to observe the system status. (The system variables observed were marked on the descriptive engine start score sheets, appendix 11.1)

Callouts (including commands) are positively defined as something that must be said, given that there is correspondence between the system status and the correct callout phrase. There is no formal callout for the instances when there is a mismatch, i.e. the system status is different from the premises that must be met before the callout can be executed. This means that trainees are free to report abnormalities anyway they like (including an altered version of the correct callout), but they should not execute the correct callout phrase (in the checklist) if the premises are not met. Further more, the checklist does not forbid verbal behaviour that is not defined as a formal callout: during the pre start checklist, trainees' often callout the checklist items. However, there is no requirement to do so (the R pilot could actually be alone in the cockpit at this point in time), but some pilots may find it helpful to memorise the checklist by verbalising certain items. Thus, repetitions of callouts, various *callout-like* statements and general small talk are therefore not categorised as errors in this analysis. Everything that is said outside a callout is irrelevant, except if an additional statement refers directly to a callout and changes the meaning reference so that the callout becomes ambiguous. For example, if the call out was “starting right” and this callout was repeated later, this is not erroneous as long as the right engine was started. If the repetition of the call out (and *callout-like statements*) was incorrect, e.g. “starting left”, then this repetition makes the meaning reference of the first callout ambiguous, hence an error of commission.

Role errors

Role errors relate to actions that are correct when seen in isolation, but performed by the wrong crewmember. There is a role reversal error when the crewmembers perform the checklist items of each other. This type of error can be continuos; only one error for each role reversal between the two trainees was scored. Role reversal is not directly related to a/c system consequences and it was scored independently of the other error types (configuration and callout) and the error modes.

General Error Modes

The word *mode* only implies that the error category is more general than the domain specific *type* of error. The domain specific errors related to *cockpit configuration* (non-verbal, including checking onboard equipment) and *callout* (verbal, including commands) can be placed in either of these *modes*:

- Omission: not doing something that should have been done
- Commission: doing something that should not have been done
- Flow: doing something correctly, in the wrong sequence/at the wrong time.

Omission and Commission

An erroneous action can in some cases be classified both as the presence of the wrong action and the absence of the correct action. However, these are two perspectives on the same error and not two independent errors. Omission is the total absence of an action related to the a/c system (within the sample). Commission is the presence of a wrong action related to the a/c system (within the sample). The presence of an action excludes the possibility of omission for that same action, regardless of how many sub systems that one action is referring to. An important aspect is to be able to isolate errors when an action was there, but something was wrong with the action, and when the action was not there at all.

When parts of a main system are controlled by separate switches, then each switch action can constitute one error; e.g. if both L and R switch should be off, then not switching off L side is one error of omission and switching R side on is one error of commission. No speculation is made regarding cognitive aspects like the trainees’ understanding of a/c systems, e.g. *if the switches should be on or off / i.e. would be only one error related to knowledge.*

If each sub system (part of a main system) is controlled by one action, the actions will be scored independently. If one switch controls two identical systems (L and R side), then connecting the wrong system and not connecting the correct system by the same action is one error: e.g. starting R engine when the callout was L engine is counted as one error. There is one *commission error* for each action deviating from the correct action. *Omission error* can only be assigned to actions that should have been in the sample, one for each missing action. Thus, omission and commission are mutually exclusive for the same action.

Omission transition:

Generally, the current engine start must be completed before initiating a new engine start. If a new engine start is initiated without completing the current (i.e. previous), then omission errors will be scored for checklist items that were not completed. Attempts to complete the previous checklist items during the new engine start will not be scored as errors of commission in the new start. Such actions will be disregarded in the new engine start. This is the main rule and omission errors scored due to this transition will be counted as any other error of omission. An exception from this had to be made for transitions between objectively defined (a priori/instructor) abnormal starts and the next start trial.

A chunk of omission errors that relate to abnormal starts were separated from the other omissions because it was suspected that this could represent a systematic error. Often the trainee continued directly from motoring due to an abnormal start (without completing the abnormal checklist) to a standard motoring start in a new engine start trial (e.g. calling out "timing" instead of starting at the first checklist item). This is one error of omission for each checklist item that has not been performed in both start trails (it is also a violation of the 70 seconds starter sequence limit and the required 3 minutes delay between starts). There are time limits both for motoring due to an abnormal start (minimum 10 seconds) and for the motoring start (maximum 30 seconds), but it is not possible (at least not reliable) to divide the total time the trainee is holding the starter switch between the two start trials. Therefore *timing* will be disregarded in these instances. During these transitions between engine starts, one configuration error will be registered for not letting go of the starter switch, one configuration error for not reading the abnormal checklist (two such errors if it is a hung start, because then it is also required to check the voltage) and finally four call out errors for not performing the first four callouts prior to motoring start. These errors will all be in the *omission transition mode*.

Note that these transitions are different from an erroneous shutdown due to the trainees' wrong interpretation of the system status: omitting to complete the last part and the first part of the checklist during transitions between starts are counted as one error for each missing action (omission transition between engine starts), whereas an erroneous shutdown induced by the trainee (with consequential new actions and error potentials) is counted as one error only (because it is by definition the same engine start, see paragraph 2.2.2).

<u>*Flow*</u>

A *flow error mode* is different from *omission error mode* in that it is a correct action, but it is in a wrong place (in the checklist)/at the wrong time/outside a defined time window, and still within the sample. There is one *flow error* for each checklist main heading performed in the wrong place. Thus, within a main heading it was only the end result that was scored, except for checking voltage before connecting a power source. *Flow errors* can be counted in several ways; taking into account e.g.: the size of the jump/number of main headings, if the actions are part of a chunk/several sub-headings, how the various actions are combined/checklist position number etc. In this evaluation it was decided to give one *flow error* per backward jump in the checklist. This operational definition may not describe all instances of flow errors correctly, but it made an otherwise very complicated scoring easier and therefore more reliable. Each time the trainee went backwards in the checklist to perform an action that should have been performed earlier (i.e. he jumped forward in the checklist at that time), it was counted as one flow error. This was regardless of the size of the jump or any other complicating factor. When two main-heading actions directly following each other switched places, the *flow error* mode was assigned the first of those actions, i.e. highest up in the checklist.

Flow-timing; during engine starts

Not complying with critical minimum and maximum time limits is erroneous and should immediately induce a set of corrective actions. The actions that must be timed are correct seen in isolation, but exceeding an absolute time limit (e.g. maximum 2 seconds *between condition lever to start* and *ignition*) may affect the a/c system (it could damage the engine). The correct action has been completed outside the given time window, i.e. at a time (place in the checklist) when the specific action was wrong. This was counted as one error per violated time limit.

Counting the Errors

All errors were considered to be of equal importance when errors were counted in this evaluation. This is because a serious aspect of these errors, i.e. a threat to sufficient safety, is the deviation from the checklist itself: the checklist is the main tool for guaranteeing a reliable and safe cockpit configuration at all times. The main principle for error categorisation is therefore to assign one error for each observed deviation from the extended checklist. This principle also applies to errors that are continuous (role reversal), errors that have a new set of actions and error potentials as a consequence (erroneous shutdown) and erroneous single actions that have more than one reference to the a/c system (left and right sides of the systems). However, when one system has several subsystems and each sub system requires one action, then each switch action is scored individually.

The main principle for counting errors was that there could only be *one error type and one error mode per action*. However, there were two exceptions: 1) *role error* was scored independently of the errors related to the a/c systems. Thus, it was possible to have [one *configuration error* OR one *callout error*] AND [one *role error*] for the same action. 2) a callout was erroneous if the checklist position of the callout was incorrect, i.e. one *flow error mode* AND if the meaning reference was incorrect, i.e. one *commission error mode* (due to phraseology and/or untrue a/c system conditions). Thus it was possible to have both a flow- and a commission error mode for the same callout.

Overview of the Objective Performance Measures:

See the error categorisation score sheet, appendix 11.3

Time:
Maximum 15 minutes to complete the formal a/c pre start checkout.

Error Types:
1. Configuration errors: positions of switches and throttles/controls in relation to the checklist and the system status.
2. Callout errors: phraseology in relation to the checklist and the system status.
3. Role errors: action performed by the wrong crewmember.

Error Modes (for configuration and call out errors):
1. Flow: too early or too late in the sequence – timing is not a critical factor.
2. Flow-timing: a critical procedural time limit is violated, i.e. the action lies outside the given time window.
3. Omission: the action is not in the sample.
4. Omission-transition: chunks of omissions (abnormal starts/motoring)
4. Commission: there is an action, but it is wrong (incomplete/inaccurate) in relation to the checklist and the system status.

4 Result

4.1 Pre Start Checkout in the Aircraft

The formal pre start checkout took place in the aircraft (a/c) cockpit and it ended the *technical training* (i.e. pre start training conducted before starting the FFS training). This checkout was the fifth time in the a/c for the CTR group (pass 5) and the first time in the a/c for the EXP group (their fifth pre start pass, i.e. pass 5). The Skyways chief instructor must *officially* approve the performance before the trainees are allowed to continue the Saab 340 training in the FFS. The checkout was approved by Skyways if the number of errors was sufficiently low, as judged subjectively by the instructors, and if the total time on the complete pre start checklist (the checkout sample) did not exceed 15 minutes.

4.1.1 Objective Performance Measures

- THE CONTROL GROUP: All 10 trainees in the CTR group passed (approved by Skyways instructors) the formal checkout in their first attempt, i.e. in pass number 5 (the first checkout pass). The mean number of errors was 2.5 errors and the mean time to complete the checklist was 11.7 minutes.

- THE EXPERIMENTAL GROUP: Two trainees in the EXP group passed the checkout in their first attempt, i.e. in pass number 5. This was without familiarisation with the cockpit; i.e. it was the first time the EXP group saw a real Saab 340 cockpit. (The EXP group was also denied travelling in the cockpit *jump seat* when on leave). Six trainees out of the remaining 8 trainees passed the checkout in their second attempt (pass 6) and the 2 remaining trainees passed in their third attempt (pass 7). This final formal pass was not used for comparison (too few trainees. The mean number of errors in pass 7 was 1.5 errors and their mean time to complete the checkout was 11 minutes).

- ADDITIONAL FAMILIARISATION PASS (8): After pass 7, the instructors decided (the 3 instructors agreed) that all 10 EXP trainees would have to participate in one additional pre start checklist *familiarisation pass* (pass 8), before they could be allowed to train in the FFS. The purpose was to make the trainees more secure on the items they performed less well and generally to enable a faster completion of the pre start checklist. The training was recorded on audio tape and the items that, according to the instructors, required additional attention in the familiarisation pass were:

1) A repetition of the pre start checklist items, mostly the practical performance rather than the theory.

2) An emphasis on how to move switches and controls in a *correct* way. This is not described in the checklist. There is actually no formal *objectively defined wrong way* of moving switches (except for the battery switch that must not be placed in *override*).

3) Items not trained by using the MATE touch screen, i.e. thought verbally and by the help of demo object/pictures/paper tiger parts of MATE (such as e.g. MEL/the torch, the oxygen mask, trim settings).

The additional pass was also used to catch up with items not trained in MATE at all, e.g. the *hands on training* (a separate type of training that deals with other onboard systems, like e.g. the manual hydraulic pump). This training had to be delayed in order to avoid cockpit familiarisation before the a/c pre start checkout (pass 5) for the EXP group (as we measured the transfer effects from a real to a virtual environment.)

The following sections will examine errors and time during the pre start checklist. The non-parametric Mann-Whitney Test was used (.05 convention).

Differences in the total number of errors

The figure below shows the comparison of errors:

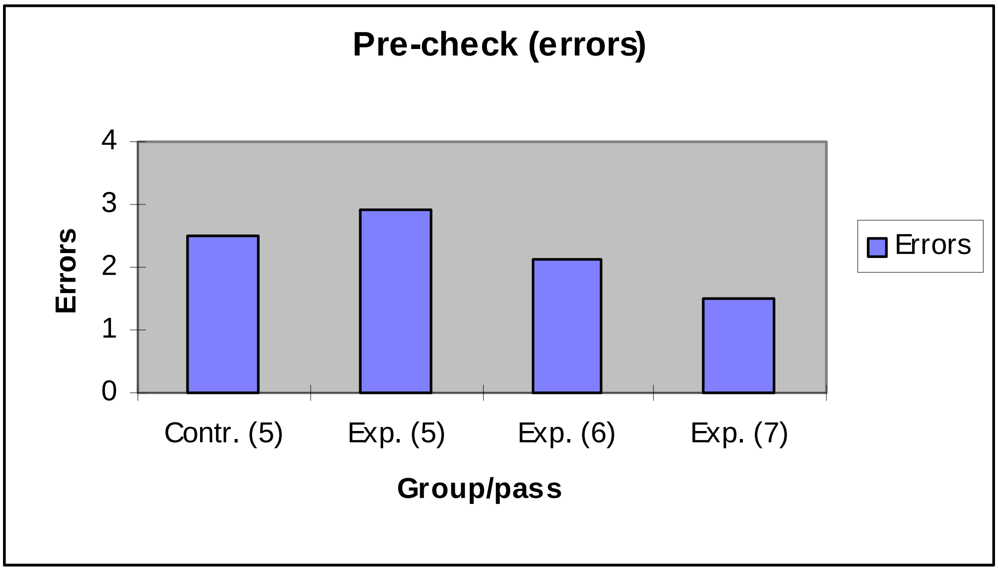


Figure 7: The number of errors (all error categories) observed during the pre start checkout in the a/c. The control group (CTR 5) was compared with the experimental group the first (EXP5) and the second (EXP 6) time the experimental group performed the checkout. (EXP 7 contained two trainees only). There were no significant differences in the mean number of errors for any of these comparisons.

1. There was no significant difference with respect to the mean number of errors observed in the CTR group checkout (pass 5) and the EXP group checkout (pass 5). This comparison was made with the inserted error potential and for the whole checkout sample, i.e. including checklist items that were not relevant to the touch screen technology in MATE (the training environment prototype included more training facilities than just the two touch screens, e.g. paper tiger parts and demo objects). The mean number of errors for the CTR group was 2.5 errors. The mean number of errors for the EXP group was 2.7 errors.

2. There was no significant difference with respect to the mean number of errors observed in the CTR group (pass 5) and the second checkout for the EXP group (pass 6). This comparison was made with the inserted error potential and for the whole checkout sample, i.e. including checklist items that were not relevant to the touch screen technology in MATE. The mean number of errors for the CTR group (pass 5) was 2.5 errors. The mean number of errors for the EXP group in pass 6 was 2.25 errors.

Differences in the time used to complete the checklist

The figure below shows the comparison of the time to complete the pre start checklist:

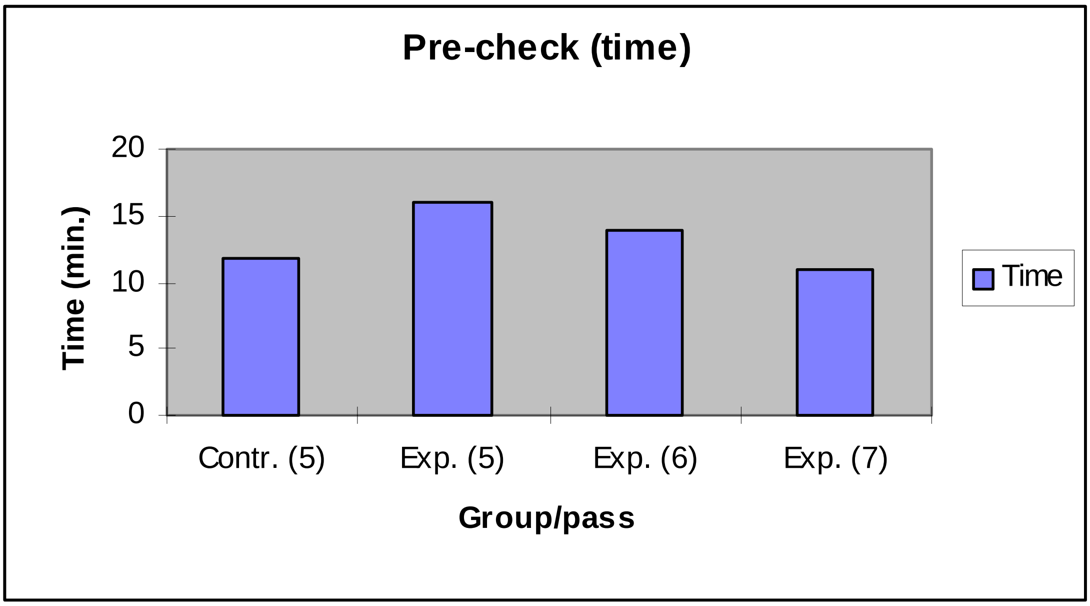


Figure 8: The mean time (in minutes) trainees used to complete the pre start checkout in the a/c. The control group (CTR 5) was compared with the experimental group the first (EXP 5) and the second (EXP 6) time the experimental group performed the checkout. (EXP 7 contains two trainees only). The experimental group (MATE trainees) used significantly longer time both the first and the second time in the a/c.

3. The EXP group used significantly longer time than the CTR group in their first checkout attempt ($p < 0.001$). The control group mean time was 11.7 minutes to complete the checkout (pass 5). In the first attempt (pass 5), the EXP group mean time to complete the checklist was 16.1 minutes. No trainee used more than 20 minutes.

4. In the second attempt for the EXP group (pass 6), the mean time to complete the checkout was14 minutes. All EXP trainees used less than 15 minutes (the official limit for approval), but still significantly ($p = 0.001$) longer mean time than the CTR group in the checkout pass 5 (11.7 minutes).

Differences in errors and time for parts of the checkout sample

The main result for errors and time for the pre start checkout sample was that the groups did not differ with respect to the amount of errors, but the EXP group was significantly slower to perform the checkout than the CTR group. An explanation for this could be that the trainees used more time on checklist items that were not simulated by using the touch screen technology in MATE, i.e. items that were *mentally* rehearsed. Items outside the training sample were explained verbally to the trainees, e.g. items on MEL (minimum equipment list). Checklist items on other cockpit panels were explained by the help of the paper tiger parts of MATE (pedestal and side panels). Finally, a real oxygen mask was shown (but not demonstrated). Thus, it could be that the extra time was spent on these items, and not on the items that were trained by using the touch screen. The graphical representations of switches etc. look quite similar to the real switches. Perhaps trainees would be unaffected by the change of environment to a real cockpit for this particular part of the sample, but slower on all the more *theoretical* items?

To test this hypothesis, the (complete) checkout sample was divided into the training part of the sample, which is relevant to the touch screen technology only, and the remaining part of the checklist sample (i.e. the whole checkout sample minus the training part of the checkout sample). The number of errors in each sample was counted, including the error potential. Outside the training sample, the presence of batteries in flashlights was used as error potential (Emergency and safety equipment, MEL). There was no video recording of the sample outside the training sample, but the instructors made notes including the trainees' total time for completing the checklist. The time spent on each part of the sample was measured for all trainees. A between comparison was made for the mean number of errors and the mean time per part of the sample:

1. There were no significant differences between the two groups with respect to the mean number of errors in the training sample during the checkout, neither in the first (pass 5) nor in the second (pass 6) checkout attempt of the experimental group.

2. There were no significant differences between the two groups with respect to the mean number of errors observed in the remaining part of the checkout sample, i.e. the part outside the touch screen relevant items, in neither of the two checkout attempts.

3. There were significant differences between the two groups with respect to the time used to complete the training sample during the first checkout (pass 5): the EXP group used significantly ($p = 0.001$) longer time than the CTR group on this touch screen part of the checkout sample.

4. There was a significant difference between the two groups with respect to the mean time used to complete the training sample during the second checkout pass (pass 6). The EXP group still used significantly ($p = 0.033$) longer time than the CTR group on this touch screen relevant part of the checkout sample.

5. There was a significant difference between the EXP group and the CTR group on the remaining part of the checkout sample, i.e. the items not relevant to the touch screen in MATE. The EXP group used significantly longer time to complete this remaining part of the checkout sample, both in their first / pass 5 ($p < 0.001$) and in their second / pass 6 ($p = 0,001$) checkout attempt.

The comparisons of objective data from the pre start checkout in the a/c indicated that the EXP group did not differ from the CTR group with respect to number of errors, but that the EXP group used significantly longer time to complete the checkout. The EXP group used significantly longer time on both the training sample (touch screen relevant only) and the remaining part of the checkout sample.

Instructors have also rated the performance subjectively on *instructor score sheets*. These ratings will be reviewed in the following.

4.1.2 Subjective Performance Measures

Instructor Score Sheets

Instructors were asked to rate the quality of the performance of each trainee from 1 (bad) to 5 (good) on the following 4 dimensions: knowledge of cockpit layout, speed in performing the checklist items, accuracy in the performance and inference of the trainees' understanding of a/c system interactions. In addition, instructors were asked to predict trainees' performance on the pre start FFS. These ratings were done immediately after each session. As shown in the figure below, the EXP group was rated lower than the CTR group on all these dimensions:

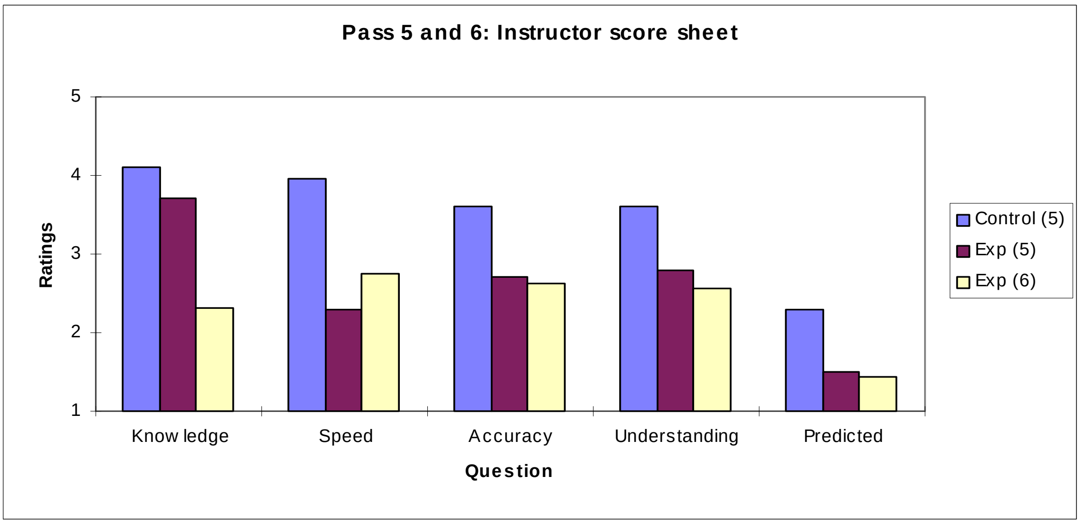


Figure 9: *Instructor's subjective ratings of the trainees' performance (5 was best). The experimental group was rated lower (worse) than the control group on all dimensions, both in their first (EXP 5) and second (EXP 6) checkout attempt. During pass 5, Speed was rated significantly lower (p = 0.002), thus in accordance with the objective data (time). The difference in Accuracy was not significant in pass 5, thus also in accordance with the objective data (errors). However, all dimensions in the second attempt (pass 6) were rated significantly lower for the experimental group.(Instructors were asked to rate the trainees according to their current stage of training, i.e. the reference changed from pass 5 to pass 6.)*

The results from the instructor score sheets during the first checkout pass (pass 5) can be summarised the following way:

- Instructors rated trainees' *knowledge* of cockpit layout lower for the EXP group than for the CTR group, but not significantly.
- The EXP group was rated significantly lower ($p = 0.002$) than the CTR group on the dimension *speed*, corresponding to the objective measure of total time to complete the checkout.
- The EXP group was rated lower than the CTR group on the dimension *accuracy*, corresponding to the objective measure of total number of errors, but not significantly.
- The EXP group was rated lower than the CTR group on the dimension a/c systems *understanding*, but not significantly.
- Instructors were also asked to predict trainees' performance (errors and time) during the pre start FFS passes (i.e. after the formal a/c checkouts). This was done as a control for possible inaccurate trainee evaluation. Instructors *predicted* that the EXP group would perform significantly worse ($p = 0.019$) than the CTR group during the three first FFS pre start passes (as rated by the FFS instructors).

Pass 6:

All dimensions in the second attempt (pass 6) were rated significantly lower for the experimental group (Knowledge, $p < 0.001$; Speed, $p < 0.001$; Accuracy, $p = 0.043$; Understanding, $p = 0.003$; Predicted, $p = 0.006$). Instructors were asked to rate the trainees according to their current stage of training, i.e. the reference changed from pass 5 to pass 6.

4.1.3 Objective and Subjective Measures

The instructors rated the EXP group as lower than the CTR group on all dimensions. However, the objective data seemed not to support this. Thus, there seemed to be a mismatch between *subjective instructor scoring* and *objective scoring* of behaviour from video. There are several possible explanations for this. Are the EXP group *slow learners* to start with, or could the three technical training instructors be biased in the favour of the CTR group? Only one of these instructors rated the EXP group in MATE. This instructor was recently certified as an instructor for technical training and he had actually been a trainee in the CTR group, trained by the main a/c instructor. The third a/c instructor was the chief instructor.

To seek answers regarding the proficiency skills of the EXP group, the objective video data during training (pass 2-4) in the two groups were compared. Data from these passes may reveal whether the CTR group prior to the evaluation was more skilled than the EXP group. Only the number of errors was considered, because the training passes were fixed to twenty minutes and each pass contained several interrupting explanations from the instructor.

Due to practical circumstances, the error potential was sometimes not inserted, or inserted in a slightly different way than described in the design (e.g. interruptions from a/c maintenance, new instructors, easy to forget the error potential that had to be inserted manually in all conditions etc.). This contributed to creating the unequal conditions for the trainees within passes. One way of looking into possible consequences of the uncontrolled error potential was to compare the total number of errors for the two groups with and without the inserted error potential.

Although the total number of errors was greatly reduced, the total number of errors observed in the EXP group did not seem to be significantly different from the total number of errors in the CTR group. Furthermore, without the error potential, the number of errors was close to zero for most of the trainees in both groups. Consequently, the error potential was needed in order to observe errors at all. The comparisons of passes 2-4 will therefore include the inserted error potential.

Below is summarised the results with respect to subjective and objective performance measures in the training passes 2 - 4:

1. There was no significant difference between the CTR group and the EXP group with respect to the mean number of errors observed in the EXP group during training in MATE (pass 2-4) and the CTR group during training in the a/c (pass 2-4).

2. The EXP group was rated significantly (pass 2; $p < 0.001$, pass 3; $p = 0.007$, pass 4; $p < 0,001$) lower than the CTR group on the knowledge of layout dimension in all training passes (2-4), even from the start of the training (pass 2).

3. A strong correlation between all dimensions was found (Spearman r, p values close to 0.001 for many of the correlation), possibly suggesting <u>one</u> *uncertainty* dimension. (The difference between subjective and objective measures could be due to a HALO effect.)

4.1.4 Summary Pre Start in the Aircraft

The EXP group does not make significantly more errors than the CTR group, but they use significantly more time to complete the pre start checkout. The majority of the trainees are approved by Skyways the second time they are in the a/c, i.e. they have an acceptable low number of errors and they use 15 minutes or less, but they still use significantly longer time to complete the checklist than does the CTR group. This is also confirmed by the instructor score sheets (pass 5): there were no significant difference in accuracy (objective: errors) but significant difference with respect to speed (objective: time). (See the Discussion section about subjective data/possible HALO effect).

4.2 Pre Start in the Full Flight Simulator

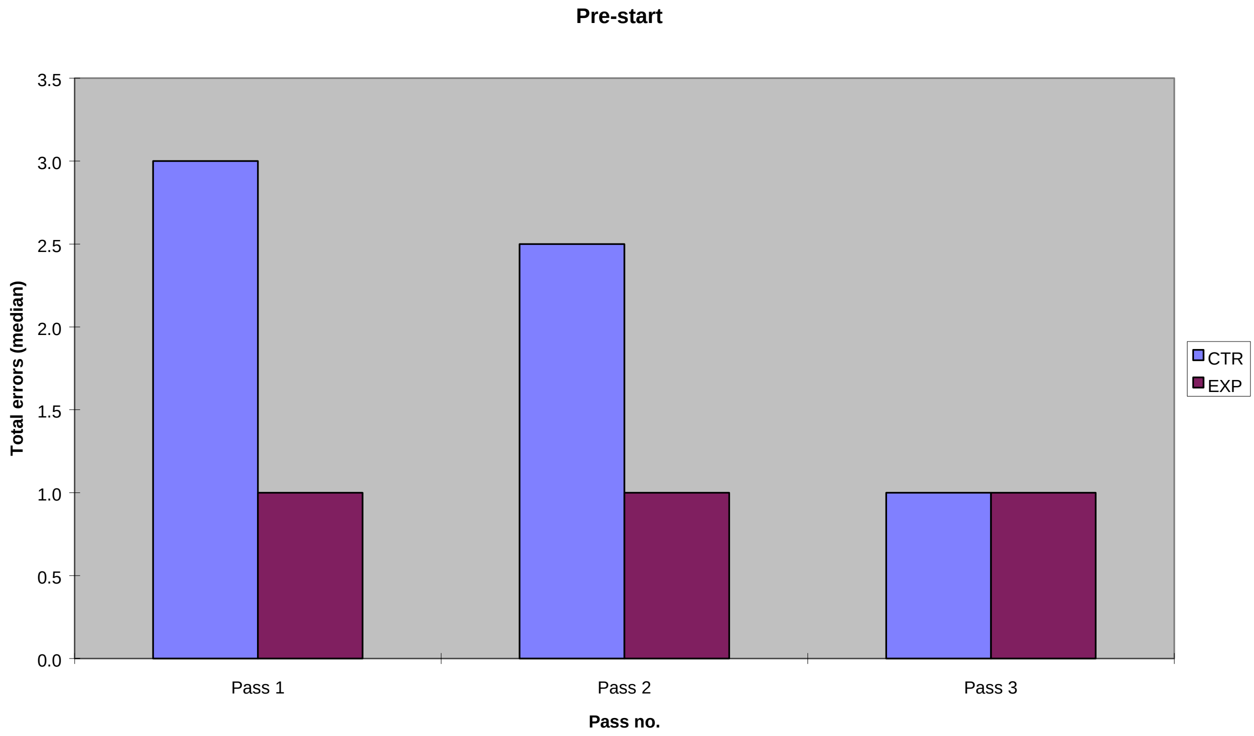


Figure 10: the number of errors (total no. error categories) in the CTR group was higher than the number of errors for the EXP (MATE) group in all three passes, but not significantly so.

1. The FFS passes were combined pre starts and engine starts. There was no systematic insertion of an error potential during the pre start parts of the passes. Hence very few errors were made and most errors were related to the difference between the A- and B-type (the auto coarsen system). When the first three FFS passes were collapsed, the CTR group had between 2.0 and 3.0 errors at an average. The EXP group had approximately 1 error in all three passes. No significant results were produced. At this point in time, the EXP group had more training in the a/c than the CTR group (familiarisation pass number 8, plus additional checkout passes).

2. The total number of errors in the CTR group was higher than the total number of errors in the EXP group in the first two passes, but not significantly. This indicates that the instructors were less good at predicting performance. The two a/c checkout instructors predicted that the EXP group would perform significantly worse than the CTR group in the first FFS passes. These predictions were made after the training passes during the a/c checkout pass 5 ($p = 0.019$) and pass 6 ($p = 0.006$). (Predictions were also made during the last training passes, but instructors were asked to predict based on the current stage of training.)

The next section will report, using the same type of tests, results from the analyses of the engine start samples in the full flight simulator. A distinction was made between normal, i.e. no inserted error potential, and abnormal engine starts, i.e. *hung* and *hot* starts (inserted abnormalities).

4.3 Normal Engine Start in the Full Flight Simulator

First, the number of errors (all categories) in both groups was compared.

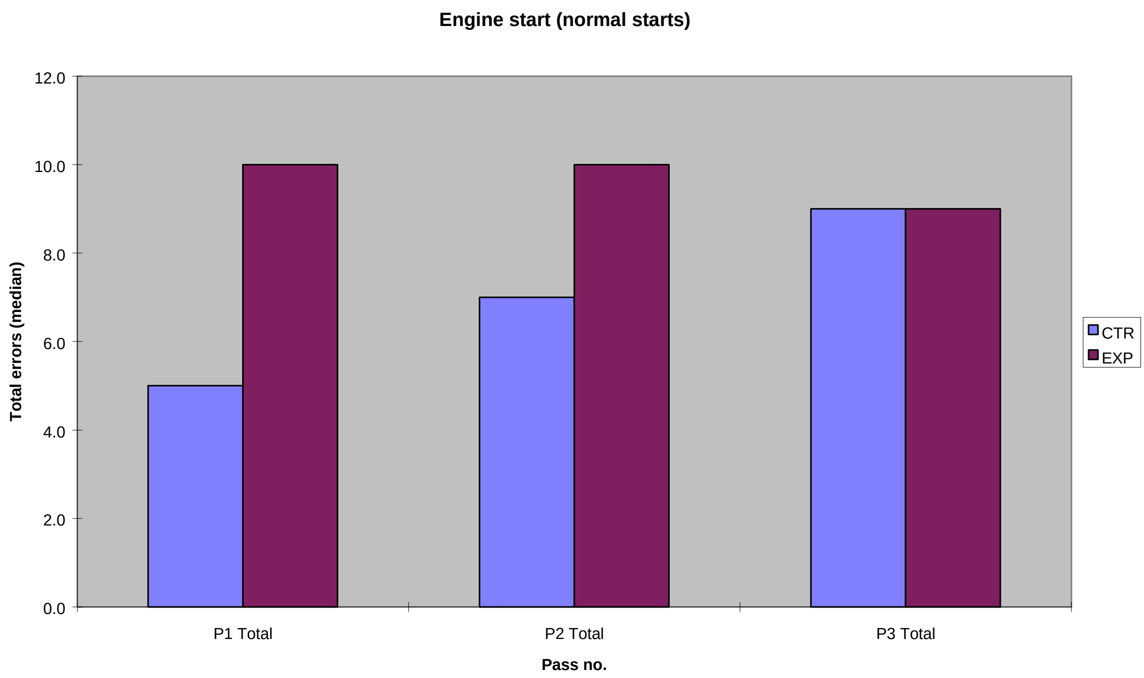


Figure 11: the number of errors (all categories).

1. When the numbers of errors (all categories) observed in the three first FFS passes were collapsed, the number of errors in the EXP group was significantly (p = 0.043) higher than in the CTR group. When looking at the three passes separately, a higher number of errors were observed in the EXP group in the two first passes, but this difference was not significant. The total number of errors was stabilised between 8 and 9 errors for both groups in pass 3.

2. As the figure below shows, callout errors stood out as the most common error during normal starts. The EXP group had more errors related to callouts than the CTR group in the first and in the second passes. This difference was significant in the first pass (p = 0.006). The CTR group had more callout errors than the EXP group in the third pass, but not significantly. When all 3 passes were collapsed, the EXP group tended towards having more callout errors than the CTR group. This tendency was very close to the .05 level (p = 0.052).

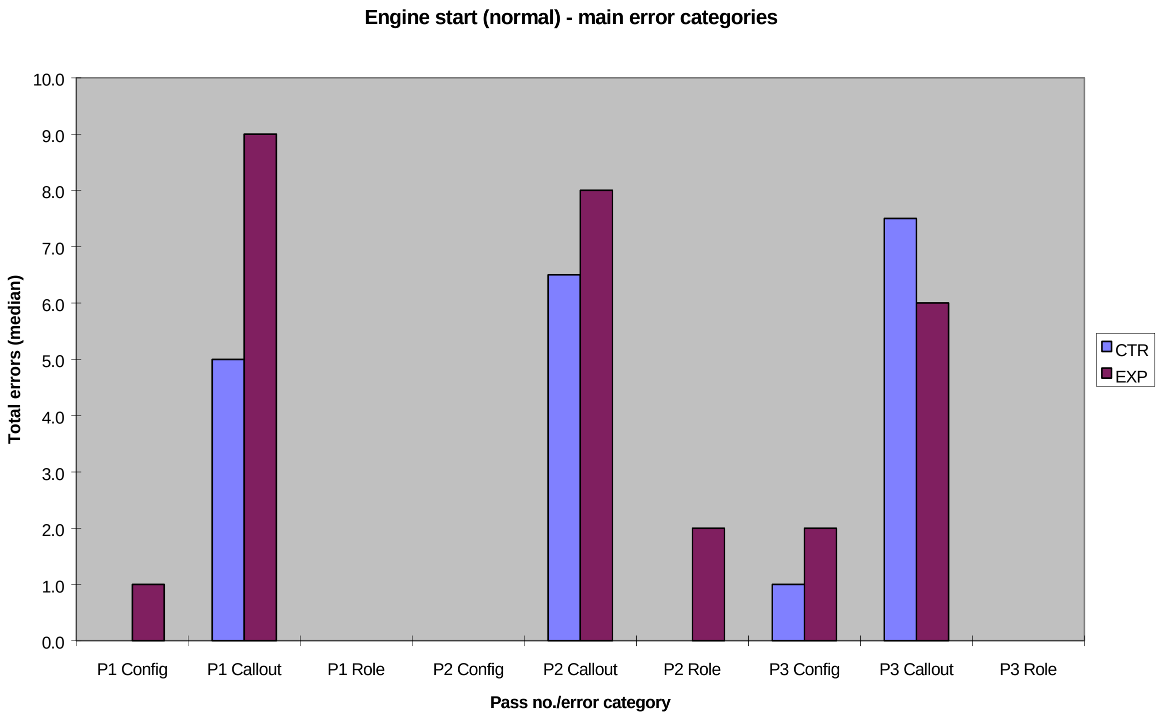
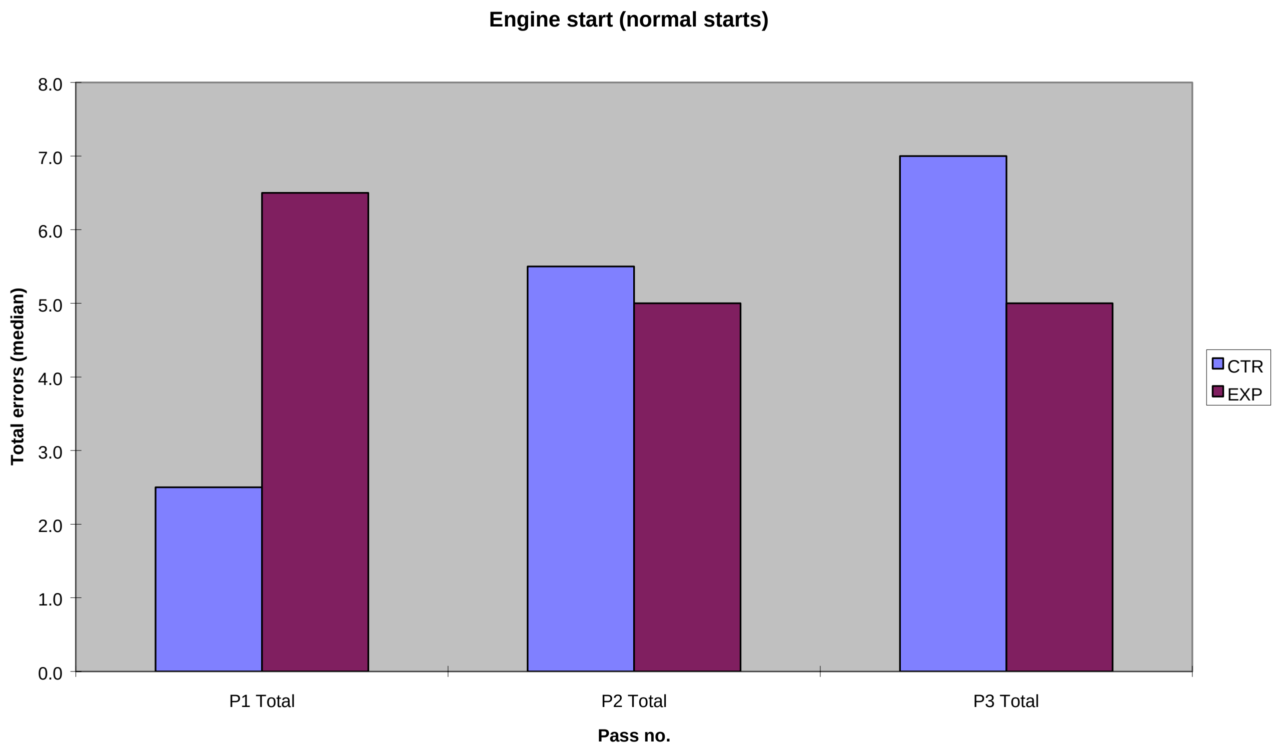


Figure 12: Callout errors stand out as the most common error in all three passes. There was a tendency for the EXP group to commit more callout errors (all three passes collapsed / average).

3. When all three passes were collapsed, there was a tendency for the EXP group to have more role errors, but there was never more than one role error per crew combination (at an average).

A closer analysis of the callout errors revealed that the majority of these were commission errors. This means that the callout phraseology deviated from the correct phraseology as described in the operational definition of the extended checklist, (but that there was a callout concerning relevant checklist items at the relevant place). If this category is removed from the data material, it turns out that:

4. There was no significant difference between the CTR and the EXP groups during normal starts when callout commission errors were removed from the comparison. See figure below.



<u>Figure 13:</u> callout commission errors were removed and no significant differences were found between the two groups during normal engine starts.

4.4 Abnormal Engine Start in the Full Flight Simulator

1. The CTR group had abnormal starts for the first time in pass 2 and the EXP group had abnormal starts their first time in pass 3. The control group had significantly ($p = 0.03$) more configuration errors than the EXP group in this comparison.

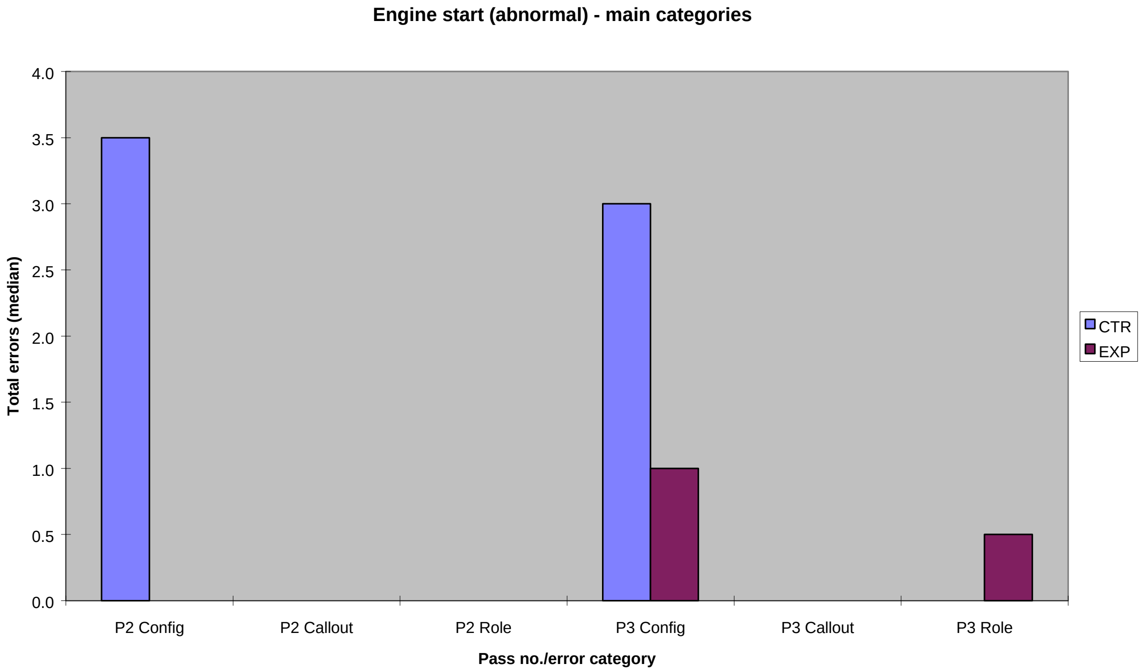


Figure 14: the CTR group had significantly more configuration errors than the EXP group. The callout commission errors have been removed.

2. When all passes with abnormal starts were collapsed (pass 2-3 for the CTR group and pass 3 for the EXP group), the number of errors observed in the CTR group was significantly ($p = 0.014$) higher than in the EXP group.

3. When all passes with abnormal starts were collapsed (average value, pass 2-3 for the CTR group and pass 3 for the EXP group), the CTR group had significantly ($p = 0.024$) more configuration errors than the EXP group.

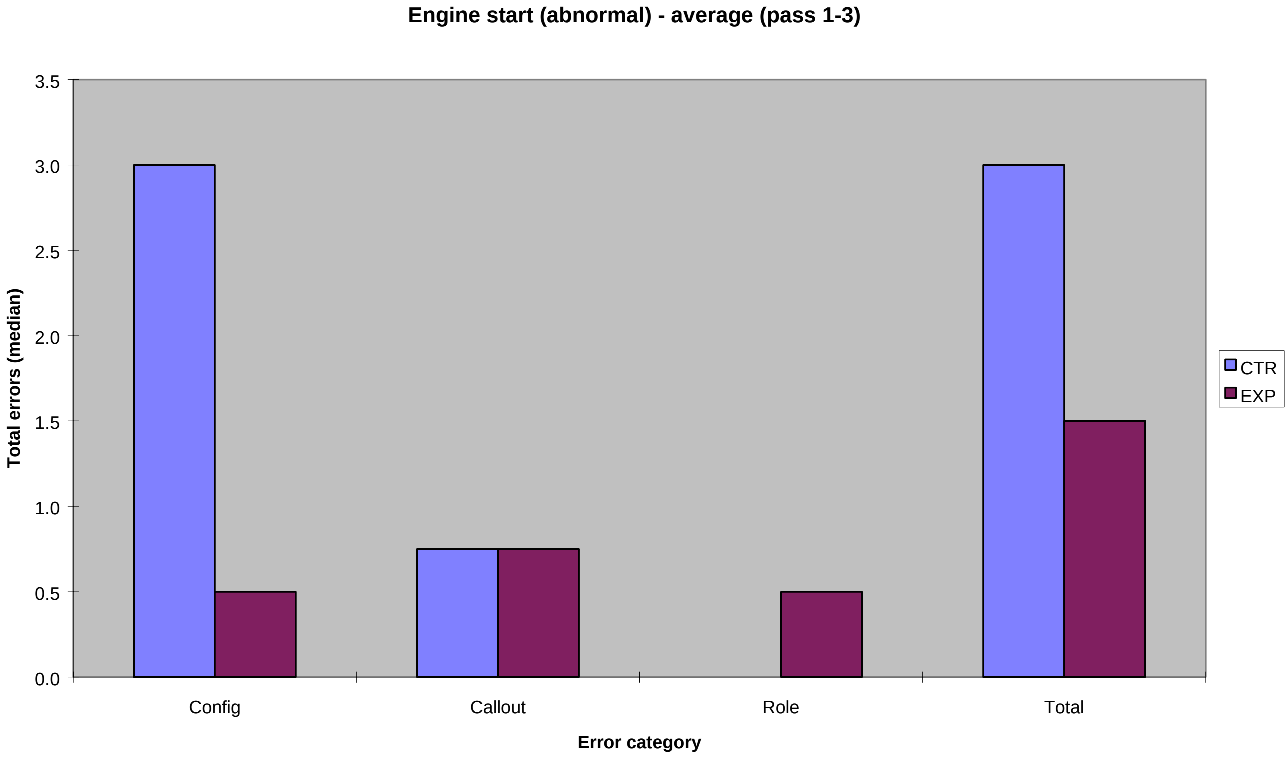


Figure 15: *number of errors when passes 2-3 were collapsed (average)*

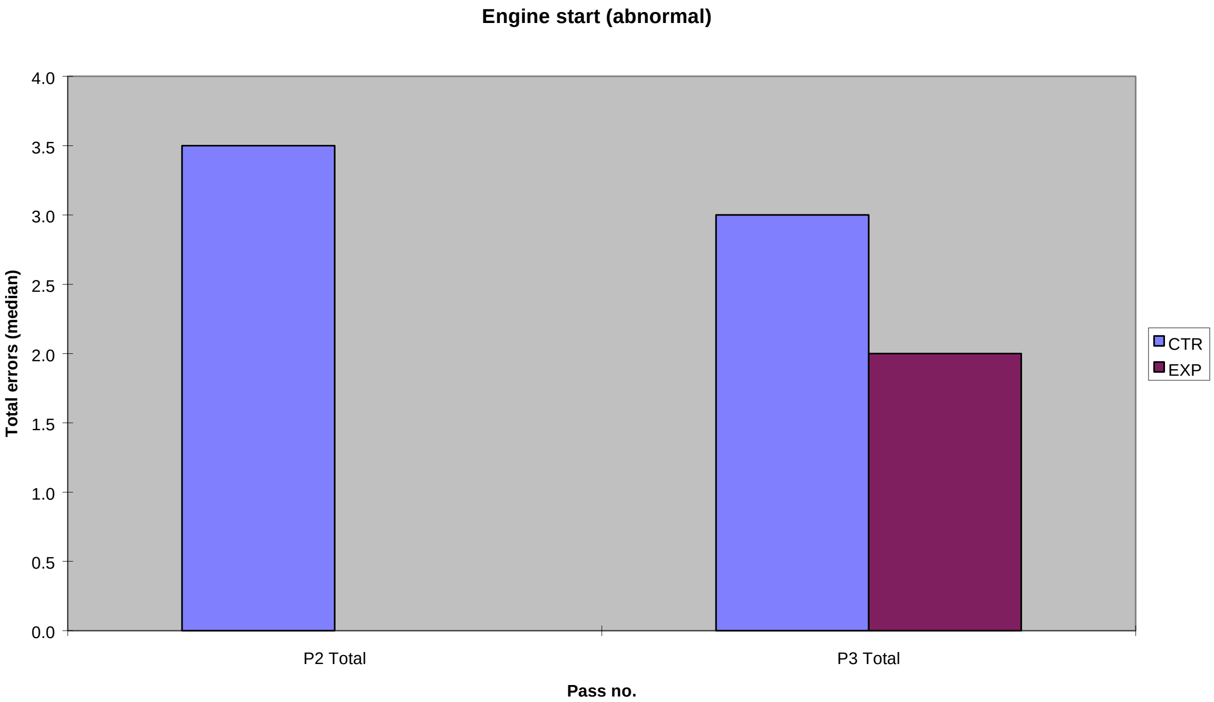


Figure 16: *the number of errors observed in the CTR group was significantly higher than the number of errors observed for the EXP group in the abnormal trials. There were no abnormal trials in the first pass. The EXP group was not exposed to abnormal trials in the second pass.*

4.5 Instructors' and Trainees' Opinions of MATE

Eight MATE trainees and 4 instructors (2 from the a/c checkout and 2 from the FFS training) have answered the opinion questionnaires (appendix 11.8). The questions were open; thus respondents were free to express any opinion they had on the MATE. The answers were interpreted, summarised in keywords and compared question by question (*item* by *item*).

<u>Negative comments:</u> these were mostly related to the pre start training. The reduced quality of performance (MATE trainees compared with the CTR group) during the a/c pre start checkout was attributed to (a) inaccurate and unreliable MATE touch screens and (b) to the lack of familiarisation with the real a/c cockpit environment, specifically non-touch screen relevant checklist items (e.g. emergency equipment). It was claimed that the MATE prototype could not replace pre start training in the a/c.

<u>Positive comments:</u> a higher level of fidelity (i.e. realism/interactivity) on training devices was appreciated on a general basis. The possibility for interaction with dynamic a/c systems was generally recognised as the strength of MATE training. It was claimed that the MATE prototype was a better engine start trainer than the power plant trainer was.

Answers that could be interpreted as in agreement with each other (1-4):

1. <u>The touch screen:</u> All 12 respondents agreed that the functionality of the touch screen in MATE was inaccurate and not reliable, and that this would have to improve considerably in order to operate switches effectively (this version of MATE had a real throttle box). The difficulties with operating switches was a continuous problem because (a) switches that were already set were moved accidentally, i.e. the touch screen was too sensitive, (b) switches that were about to be set could not be moved in a proper way, (c) it was easy to loose contact with a switch while moving it, and (d) the simultaneous operation of two switches was very difficult (dual input). The *starter* and *ignition* switches were mentioned as very problematic. These problems caused the MATE training to be unnecessary frustrating and time consuming. One of the a/c instructors thought that the touch screen problems also caused a decreased motivation for training and a general confusion regarding the real a/c systems functionality.

2. <u>Familiarisation with the a/c:</u> The 2 a/c instructors agreed that MATE trainees performed worse than the CTR group during the a/c checkout. The instructors felt that MATE trainees were *less confident* in the real cockpit, as compared to the CTR group. The CTR group was more *at home*, they *had the feel*. These aspects of the performance in the cockpit were not well defined by the instructors, but the importance of these instructor opinions were emphasised by the extra familiarisation pass assigned to the MATE (EXP) group when the checkout was completed. The MATE trainees also emphasised the need for familiarisation with the real cockpit as a part of the transfer from MATE. All respondents agreed that MATE could not stand-alone, i.e. replace the a/c training.

3. Performance in the FFS: The 2 FFS instructors agreed that no important differences between the trainees in the FFS could be attributed to MATE training. The trainees did not seem to think that difficulties in the MATE affected their performance in the FFS negatively. All respondents seemed to agree that a training environment with a relatively high level of fidelity would facilitate the training of dynamic tasks better than environments with a lower level of fidelity. Although not confirmed by the 2 FFS instructors, the MATE trainees agreed on the impression that engine start training (normal and abnormal starts) in MATE would produce better results than engine start training in the power plant trainer. This impression was attributed to the increased fidelity, i.e. realism, in MATE.

4. Ideas on how to improve the MATE: the touch screen problems must be solved. Given that MATE actually works (touch screens), the mock-up could be improved by building a back panel so that emergency equipment, manuals etc. can be included. Panel lighting (contrast) should also be improved and more sound effects should be included to increase the realism. More a/c systems should be implemented, so that MATE can be used for training procedures related to automation and teamwork (2 pilot system).

5 Discussion

Instructors were responsible for most of the practical day-to-day data recording (except in the a/c), but they received different amount of briefing with respect to recording techniques and the design. This was due to practical limitations in their schedules (they were mostly flying when not instructing), but they all received the same written instructions for how to record data. *Good will* from the Skyways personnel ensured a minimum loss of data, but attempts to include experimental control suffered from these variations (opportunity to have briefings), specifically the *default configurations* and the *inserted error potentials* during the pre-start sample. Generally, the limited possibility to control variables in the evaluation was a direct consequence of following real training courses: the first priority was to train pilots effectively and not to evaluate MATE.

The error categorisation system was sensitive to differences underlying the first impression of the data material. By combining error types and error modes, we found that (a) *callout commission error*, i.e. incorrect callout phrases, was the main factor responsible for the differences in favour of the CTR group during normal starts and (b) the CTR group had significantly more *configuration errors* during abnormal starts. These were important findings that would have been overlooked by limiting the analyses to comparing the total number of errors. However, further analyses on this level of resolution, i.e. looking at all combinations of *types* and *modes*, did not give any clear picture, except for the limited impact of *omission transition errors*. There were too few observations in each category to look closer at each combination of error type and mode (no significant results were produced).

The main evaluation of the pre start training in MATE was done in comparison with the real a/c, during the checkout of real pilots. Furthermore, the MATE prototype was only a *live paper tiger*, i.e. no special pedagogical tools were implemented. This gave little room for the MATE to stand out as a better alternative (than a/c and paper tiger) for the pre start part of the training. Further more, the main task during this checklist, i.e. checking cockpit configuration, can be classified as a relatively simple pattern recognition task. One would perhaps not expect a simulation of dynamic system interaction to be necessary in tasks that mostly require picture like representations, i.e. the paper tiger and the checklist may be sufficient for learning the cockpit layout and configuration previous to the a/c cockpit familiarisation. Generally, learning new procedures in a new a/c system is different from training (new) procedures in a well-known a/c system. One would expect a novice pilot to need as much *practical experience* with the real a/c as possible. An experienced pilot would maybe not have this need for high fidelity during training of procedures. Thus, flying by instruments, crew resource management training (CRM) and testing of new instrumentation could be interesting use of MATE.

One expectation was that the EXP group could end up having more errors than the CTR group when transferred directly to the a/c, but that the EXP group would quickly approach the acceptable level of errors (that the control group had). Instead it was found that the amount of errors did not differ between the groups, but that the EXP group used significantly longer time to complete the pre start checklist. The extra use of time is in accordance with preliminary studies (Andersen and Hansen, 1996). However, no trainee used more than 20 minutes. In their second attempt, all EXP trainees used less than 15 minutes (the official limit for approval), but still significantly longer time than the CTR group. (The mean time in this second checkout was14 minutes.) The question can be raised if the time used to complete the checklist really should be unacceptable if it was below 20 minutes. The total time for completing the checklist is not related to the immediate safe configuration of the cockpit and the Skyways 15 minutes limit seems tough compared with the 20 minutes limit at SAS.

Skyways arranged in total 3 extra a/c passes for the EXP group (checkout 6-7 and familiarisation pass 8). The CTR group had 4 training passes before the checkout pass (pass 5), thus it seems like the number of passes that can be saved in the a/c is marginal. The reduction of a/c passes in comparison with the CTR group was minimum 1 pass. The reduction was maximum 3 passes if the familiarisation pass number 8 and pass number 7 (two trainees) are excluded. Thus, the use of the MATE prototype for this type of pre start training may be sufficient in terms of errors and time, but users of the prototype set-up of MATE may not view MATE training as cost effective for training pre starts. It can also be argued that training normal engine starts is sufficient in the power plant trainer. However, MATE may provide better training than the power plant trainer for handling abnormal engine starts.

In this evaluation, less emphasis was put on the *time* to complete the pre start checklist. This was because: (1) there is no direct a/c system consequences related to the *total time*, and (2) because it seems to be practically less important, as exemplified with the different time limits in Skyways and SAS. However, Skyways instructors were of the opinion that the checklist performance had to be improved for all EXP trainees by adding a *familiarisation* pass (pass 8). Most trainees from the EXP group were officially approved by Skyways the second time in the a/c, and the question concerning the need for an additional familiarisation pass that included training on the pre start checklist samples can therefore be raised. Given that the EXP group really needed an additional pass related to these tasks, one might suspect that the approval of the EXP group was due to maintaining the group motivation, i.e. no trainee should suffer from being a member of the EXP group. Thus, a familiarisation pass would be better than a series of failures in formal checkout passes.

The instructors consistently rated the CTR group higher than the EXP group. Does this imply that the CTR group was better to start with? It was not possible to allocate the subjects randomly to the groups or to match subjects on background variables thought to be important. However, there did not seem to be differences between the groups with respect to the objectively scored number of errors during training (passes 2-4). The requirement for familiarisation could be due to a possible HALO effect, i.e. a consistent instructor bias in the disfavour of the EXP group. The two instructors (during checkout) predicted that the CTR group would perform significantly better than the EXP group on the pre start checklist when rated in the FFS. These predictions were made after the training passes during the a/c checkout pass 5 ($p = 0.019$) and pass 6 ($p = 0.006$). However, the analysis of objective data (number of errors) showed a different result, i.e. the EXP group performed slightly better than the CTR group during the pre starts checklist in the FFS, but not significantly. The EXP group now had additional familiarisation related to the pre start, but these results may indicate that the instructors were less good at predicting the performance of trainees: the ratings were always in favour of the CTR group, i.e. by definition no *top guns* in the EXP group. Further more, all the instructor dimensions correlate strongly (Spearman r, $p < 0.01$, for 6 out of 10 combinations) and all the significant correlations were found in the EXP group. Thus, it could be argued that the instructor evaluation in the EXP group was on a more *general* level, perhaps an *uncertainty* (or *confidence*) dimension. A HALO effect is also supported from the fact that EXP trainees received significantly lower ratings on *knowledge of cockpit layout* already from their <u>first</u> training pass, i.e. directly after training in paper tiger (where they had equal amount of training). Perhaps the perceived overall quality of the EXP group performance was coloured by the need for more time to complete the checklist? There is also a risk that the objective measures were insensitive to important aspects of the trainees' performance and that the instructors were able to observe other (more ill defined) aspects of how *confident* the trainees were. In the opinion questionnaires, the instructors emphasised that MATE trainees were less confident and unfamiliar with the cockpit. This was also confirmed by the trainees and it could be a reason for the relatively low ratings. It is perhaps likely that experienced instructors can sense how *confident/secure* a trainee is in the cockpit. Nevertheless, all MATE trainees had to participate in familiarisation, but the familiarisation problem was limited to the a/c checkout. When trainees were approved for FFS training, this difference had disappeared. The FFS instructors answered (opinion questionnaires) that no important differences observed between trainees in the FFS could be attributed to the MATE training (due to the project delay and limited analysis resources, the FFS instructor's subjective ratings were not included in the evaluation).

The answers from the FFS instructors is also relevant with respect to the indication that MATE trainees performed better on abnormal engine starts: FFS instructors did not mention this difference, although MATE trainees seemed to be convinced that they were good at diagnosing abnormal starts due to MATE training. These trainees have not trained abnormal starts in the power plant trainer (but they were instructed on normal engine starts in the power plant trainer and they have possibly discussed engine start training with other pilots). It is also clear that the two types of abnormal starts look very similar on the instrument readings (both starts look *hot*, i.e. high ITT) in both the power plant trainer and the MATE. Thus, an actual improvement of abnormal engine start diagnosis due to MATE training can be questioned. However, trainees were convinced that the increased level of fidelity/realism was a valuable support in the engine start training. There is support for this opinion in the objective data that indicate a better handling of abnormal engine starts for the MATE group. The analysis of the data from the engine starts (passes 1-3) showed that the total number of errors observed in the EXP group was higher than for the CTR group during normal engine starts, but almost all the errors were related to callouts. There is no obvious reason why callout errors should be strongly related to training in MATE. The tendency is in a different direction during abnormal starts. Here, almost all errors are related to configuration errors and they are mainly in the CTR group. Hence, the tendency is that the EXP group has a better performance during abnormal starts than the CTR group.

The operational definition of a correct callout included the verbatim execution of the callout phraseology given in the extended checklist. The descriptive scoring had as an objective to record the callouts word-for-word. This is a relatively easy (and probably reliable) way of scoring, i.e. to avoid judgements about what should constitute a clear callout. The scoring of callouts was successful in most instances, but sometimes inhibited by the problems with sound in the FFS trails with the EXP group. In these instances, the scoring was related to the parts of the callout possible to hear, i.e. the morphemes. In sum, the scoring and categorisation of callouts has been very detailed: the smallest observable deviation from the correct phrase in the checklist would constitute a callout error of commission. On the other hand, perhaps the criteria used for scoring callouts were <u>too</u> detailed, i.e. that some deviation from the correct callout should be allowed (if it can still be understood). Exactly how the callout deviates from the checklist was described in the score sheet, but such differences were not incorporated in the error categorisation system. To make the criteria for a correct callout less detailed, the two groups were compared without callout errors of commission, i.e. without the callouts with inaccurate phraseology (regardless how serious inaccuracy). The tendency for the EXP group to commit more callout errors than the CTR group disappeared and there were no significant differences between the groups on any categories during normal starts. There was a tendency for the EXP group to be involved in more role errors, but the number of role errors was very low in all passes (approximately 1 error at an average). When trainees uttered the callouts inaccurately, they were usually not corrected by the instructors. Neither were trainees reminded if the callout was omitted. When instructors did correct the use of a callout, they most often did not execute the callout in the exact word-for-word format described in the extended checklist. Hence, the feedback provided to the trainees did not support learning the callouts exactly the way they are described in the extended checklist. This instructor practice (at least during FFS instructions in this evaluation) demonstrates a lower emphasis on the accuracy of callouts than the correctness of cockpit configuration. Such a practice, if it exists in line operation, is to some extent conflicting with the purpose of using formalised communication.

An important part of the results is the indication that training in MATE may improve performance during abnormal engine starts. This inference is based on the difference in configuration errors. Less emphasis was placed on the fact that the EXP group had more callout commission errors during normal starts. There are several reasons for this emphasis on error types. First, instructors did not themselves put great emphasis on callout errors. Perhaps this type of errors is viewed as secondary to configuration errors. A possible explanation for this could be related to the perceived seriousness of the callout error: a callout can be wrong in relation to the operational definition, but still be interpreted correctly by trainees (and instructors). Thus, the callout error may not cause inefficient communication about the system status. Trainees were often corrected when involved with configuration errors, but rarely corrected (and inaccurately corrected) when involved with callout errors. Consequentially, callout errors stand out as a major error category, especially during normal starts. This is one reason to compare the groups without callout commission errors. Further more, the potential for committing callout errors is low if a malfunctioning start is detected (and all abnormal starts were detected), because the abnormal checklist contains more configuration actions. The abnormal start is probably also more difficult than normal starts, because more system feedback must be interpreted and corrective actions must be initiated within given time windows. Thus, the shift from callout errors during normal starts to configuration errors during abnormal starts is not surprising. The interesting part is that 1) a significantly higher amount of configuration errors was observed for the CTR group and 2) this was independent of the removal of errors in the omission transition mode (representing between 2 – 3 possible configuration errors in the abnormal checklist). Higher fidelity, i.e. spatial relationships, scanning of instruments and a more realistic environment in the MATE than in the power plant trainer stand out as a possible explanation to this effect. Further evaluations of MATE could include e.g. measures of scan patterns and cognitive measures of diagnosis to explain why MATE is supporting the training of abnormal starts better than the power plant trainer.

Omission transition errors were a sub category of omission errors during the transition from an abnormal start trial to the next engine start trial. Omission transition errors were observed only in the CTR group and they were mainly related to motoring and callouts (see the operational definition of the checklist, appendix 11.1). These errors were observed during transitions between separate engine starts involving pre-defined abnormalities, but the trainees did not complete the abnormal checklist nor did they complete the beginning of the next checklist reading. (However, if a trainee shut down for any other reason than the pre defined error potential, the repetition of checklist items up to the point where the trainee failed was disregarded, see the analysis chapter.) This specific type of omission error represented approximately 20% of all the errors observed in the CTR group during abnormal engine starts (51 errors out of 253 errors total in pass 2-3). Omission transition errors represented approximately half (49%) of all the CTR group omission errors during abnormal starts (51 omission transition errors out of 104 omission errors in pass 2-3). Thus, one would expect this type of error to have an impact on the comparison: if the omission transition errors are included in the data comparison (pass 2-3 in the FFS), then the potential for errors in the CTR group is higher than without omission transition errors. This could bias the comparison to the advantage of the EXP group. If the omission transition errors are removed from the comparison, then the potential for errors in the CTR group decreases. This potential for errors includes callout omission errors, i.e. the first four callouts of the next start trial are missing. The removal of these transition errors could bias the comparison to the disadvantage of the EXP group, especially since a tendency seemed to be that the EXP group had more callout errors than the CTR group (although these errors turned out to be callout commission errors).

Recall that the CTR group had significantly more errors in total than the EXP group during abnormal starts, both when (a) pass 2 and pass 3 were compared (first trails with abnormal starts) and when (b) these two passes were collapsed (average). The main result was unchanged when the omission transition errors were removed, but the difference in the total number of errors during abnormal starts was less vivid. When the omission transition errors were removed, i.e. a disadvantage to the EXP group, then (a) no significant difference was found in the total number of errors in the pass 2 versus pass 3 comparison and (b) the significant difference in the total number of errors (average pass 2-3) was reduced to a tendency ($p = 0.106$). However, the CTR group still had significantly more configuration errors, so the main result for abnormal starts was not changed by the comparison with and without omission transition errors.

It is unclear why it was trainees in the CTR group that had transition errors. One explanation could be that instructors allowed these jumps between incomplete start trials. However, it is not likely that the instructors told the CTR group trainees, implicitly or explicitly, to perform these shortcuts because: (a) the FFS instructors instructed in both groups (and when asked, the chief instructor confirmed that each engine start trial should be complete), (b) there was not much time to save (seconds) and (c) the handling of abnormal starts was a major part of the engine start training. It is important to be familiar with the use of the abnormal checklist, to know why motoring is needed and to have practice in the correct timing of motoring (the 3 minutes delay between starts is of course disregarded during expensive simulator time). It is also important to practice the correct callouts at the beginning of a new start trial.

Thus, it seems like the trainees in the CTR group have put less emphasis on the correct handling of abnormal engine starts. This could be partly explained by their future role in line operation: the crew starts the engine, but it is the captain (L pilot) that has the responsibility and it is the captain that performs shutdown and motoring. The first officers will not be required to do this part of engine starts in line operation. All trainees in the CTR group were to be licensed as first officers, but so were the trainees in the EXP group (one trainee in the EXP group was to be licensed as a captain). First officers must nevertheless learn how to start the engines, in order to co-operate with the captain. For training purposes, this type of errors should not be allowed because the learning of correct abnormal and callout procedures will not be optimal.

A third speculation as to why CTR group trainees had omission transition errors could be related to the fidelity of the training environments. The power plant trainer (CTR) had real a/c start and ignition switches, but the representation of the cockpit (paper tiger/overhead panel) and the a/c instruments (desktop computer) was more arbitrary than in the MATE. It could be speculated that the higher fidelity of the visual surroundings, the a/c systems feedback and the role-play in MATE might have encouraged a more realistic checklist performance. Nevertheless, the CTR group should have had feedback on this type of erroneous transitions, in order to perform correct and complete sessions with abnormal starts. There is no doubt that e.g. a hot start must be handled by immediate shutdown and motoring and shortcuts are not allowed in any checklist reading.

6 Lessons Learned

This section deals with practical/technical problems during the evaluation.

- The MATE set-up tested in this evaluation was the first prototype. The touch screen was not fully developed until the evaluation started and several of the training features thought to be superior to existing training were actually not implemented at the time of testing. This includes *the procedure training assistance* (PTA) and the *debriefing and analysis tool*. These features should be tested to see if they could actually improve training effects compared to existing training in e.g. the power plant trainer. The CTR group had problems related to transitions (abnormal starts) and the EXP group had problems related to callouts. These two error types stand out as examples of training issues to be emphasised in the PTA, when this is to be implemented in the MATE. Also, the implementation of an *electronic checklist* could enable the MATE to monitor checklist performance by means of e.g. word recognition and eye movement tracking.

- Skyways was not a partner in the project, thus the project had to rely on the *good will* of instructors and trainees. All personnel and trainees at Skyways showed interest in the project and an exceptionally *good will* to make things happen. In retrospect however, it turns out that this positive attitude was insufficient for conducting this type of data recording, e.g. maintaining a sufficient degree of experimental control. The personnel collecting data in MATE and FFS, mostly Skyways instructors, should be committed in a much stronger way in order to take more responsibility for all the problems that are inevitable during such complex data collection in the field. Key personnel could be paid and trained, or a researcher could be present during all the data collection. The cost of having one fully informed researcher/research assistant (and preferably a backup researcher) in the field during data collection is probably worth the price when measured against the data quality, delayed analyses, the cost of long distance trouble shooting and a huge amount of travelling between locations (Risø is in Denmark. Data collection took place at Arlanda in Sweden). A more permanent stay in the field would probably enable the researcher to do the time consuming backup copying and perhaps the first raw scoring of data. Also, various delays of Skyway's training schedule caused a general delay of the evaluation and some cancellations of analyses. Such unexpected delays are nobody's fault, but they must to some extent be a part of the evaluation planning.

- There were no log facilities implemented in MATE. A log could have made video analyses of the MATE cockpit configuration unnecessary (leaving time to focus on other parts of the analyses).

- Another aspect of the MATE system that was not fully implemented was *the instructor control station*. The error potential had to be set manually, thus leaving a potential for errors also when inserting this error potential. In retrospect it is clear that the potential for errors of omissions was different both within and between passes (all conditions). A more consistent pre defined default cockpit configuration could have contributed to solve this problem. Such a pre defined configuration could be automatic in MATE (but not in the other conditions). The only pre defined automatic configuration possible in MATE at the time was the malfunctioning of lamps (and this was only observed during the checkout).

- The tasks in the selected samples should have been tested in pilot studies. This would have enabled a more thorough consideration of what methods to apply and perhaps influenced on how to design the training tasks. This would have influenced the use of error potentials and perhaps additional methodologies to observe e.g. system variables, pilot's information gathering and cognitive measures like pilot's diagnosis. An example: a considerable amount of time was used to develop a measure of trainees' diagnosis of abnormal engine starts. However, it was later discovered that the training devices for engine start, i.e. the power plant trainer and the MATE, did not display vivid differences between the two types of abnormal starts used in this evaluation. MATE used the same system simulation (algorithms) as the power plant trainer. It turned out that the symptoms of a hung start were almost identical to the symptoms of a hot start: both instances looked like a hot start (abnormal ITT). An important part of training the diagnosis of these two types of abnormal starts, was to monitor <u>both</u> the Ng display (possible hung start / Ng is a measure of the gas turbine acceleration) and the ITT display (possible hot start / ITT is a measure of the temperature between the gas and power turbine). According to an experienced instructor at SAS Flight Academy, it would be almost impossible for a novice to spot the difference between hung and hot starts, given the displayed symptoms in the power plant trainer and in the MATE. Trainees would be able to diagnose the start as abnormal just from monitoring the ITT display (i.e. a relatively easy task compared to monitor both displays), but less able to elaborate on what kind of abnormality this indication represented. It is possible to have hung starts developing into hot starts (dependent of the cause of the hung start), but these instances are perhaps not good examples for learning about the main differences between hung and hot starts. In the choice of making MATE a better training device (i.e. making this difference more vivid), versus the control with equal training conditions, it was decided not to improve the simulation of differences between hung and hot starts in MATE. Thus the measure of diagnosis had to be taken out of the evaluation. This could perhaps have been avoided if various pre studies/investigations of the tasks had been performed, so that the appropriate measures could be developed before the evaluation started.

Despite all this, the loss of data was not serious and we have gained more experience with data collection in the field, i.e. in a realistic environment, without any kind of support for data collection (e.g. no system logs in any of the conditions, portable battery power only, no separate lighting for video, very limited working space, not allowed to interfere etc.).

7 Conclusion

Summary of Results:
It was clear that the MATE touch screen must be improved. The main findings concerning training effects were: (1) the amount of errors observed in the experimental MATE group did not differ from the amount of errors observed in the control group, as far as the pre-start checklist was concerned. However, the MATE trainees (EXP) did use significantly longer time to complete the pre start tasks and the instructors perceived the MATE trainees as less *confident*. MATE trainees need familiarisation with the a/c after pre start training in the MATE prototype. This familiarisation has at least two aspects: (1) the touch screen simulation of real electro mechanical switches and controls and (2) managing tasks that were explained using paper tiger parts of MATE (pictures) and demo objects. Less a/c familiarisation was needed than the total amount of a/c passes needed to train and check out the CTR group. Thus, the amount of (expensive) time needed for training in the aircraft was slightly reduced. (2) MATE trainees seemed to have more errors related to callouts than the control group during normal engine starts, but a more thorough analysis indicated that the two groups performed equally well on normal engine starts. (3) The control group had significantly more errors related to cockpit configuration than the MATE trainees had during abnormal engine starts. Thus, the MATE trainees performed better than the control group during abnormal engine starts.

Main Conclusion:
The comparison of MATE trainees with conventional trainees revealed that when MATE trainees transferred to the real a/c without any previous a/c familiarisation, they used slightly longer *time* to complete the pre start checklist. The MATE trainees were not involved in more checklist *errors* than the conventional trainees were, but the MATE trainees seemed to be *less confident* than conventional trainees were. When the MATE trainees had become familiar with the real cockpit, then no important differences could be observed between the two groups, except for performance during abnormal starts. MATE trainees had better performance on abnormal engine starts. The main conclusion regarding training effects is that (a) the performance of MATE trainees is good enough if they are given *familiarisation* with the real cockpit environment and (b) the major value of the MATE prototype seems to lie in the training of *dynamic* tasks, such as engine starts.

8 Abbreviations

a/c	aircraft
AV	Audio/Visual
CTR	Control group
DERA	Defence and Evaluation Research Agency, UK
DVD	Digital Video
EXP	Experimental group
FFS	Full Flight Simulator
L	Left (pilot)
MATE	Multi Aircraft Training Environment
MEL	Minimum Equipment List (part of the checklist)
R	Right (pilot)
RISØ	Risø National Laboratory, Denmark
SAAB 340	Commuter aircraft used by Skyways (A-type and B-type)
Skyways	Swedish airline
SVHS	Super VHS
SVHS-C	Super CVHS-Compact
VHS	Ordinary (black and white)

9 References

Andersen H.B. and Hansen J.P. (1996): Comparing transfer of training of a standard control panel and a touch screen panel. Risø National Laboratory, Risø-R-917.

Rasmussen, J. (1982): Human Errors. A taxonomy for describing human malfunction in industrial installations. Journal of Occupational Accidents, 4 (2-4), 311-35.

Reason, J. (1990): Human Error. Cambridge University Press.

MATE deliverables D211, D214 and D231

10 Appendix

Instructions to instructors/subjects and tools for measurements are included. Font size and layout are edited to save space. Detailed instructions on how to operate equipment, video logs etc. are left out.

Contents:

10.1 Checklist with Score Sheet

This is the operational definition (version 6.0) of the Skyways extended checklist (version 960901) used for describing deviating actions, one complete sheet for each session.

CTR-FFS-pre (1-3) + eng (1-3) sample

Instructor:	**Pass No:**	**Crew Letter LP:**	**Crew Letter RP:**

Recording Date:	Score 1 date: By:	Score 2 date: By: Average time spent:
Data File Name: en _(**M**ate/**F**FS)_(pass no.)(2 x unique crew letters)		Data description: CTR-FFS-pre(1-2) + eng (1-6) sample

L Start No. 1 (comment):	Type Normal:	Type Abnormal:
L Start No. 2 (comment):	Type Normal:	Type Abnormal:

R Start No. 1 (comment):	Type Normal:	Type Abnormal:
R Start No. 2 (comment):	Type Normal:	Type Abnormal:

Video “overhead panel” / Tape Number (D original):	Event:	Counter:
Video “throttlebox” / Tape Number (VHS original):	Event:	Counter:

Observable pilot performance (actions) deviating from the performance described in the Skyways extended checklist 960901:

Template for pre start scoring in FFS is included. The pre start error potential is logged in the instructor pre + eng score sheet.

Engine start sample: from “Start of Engines” (engine checklist item no. 1) to “After Engine Start”. An “engine start” (normal and abnormal) is the attempt of one crew combination to start one of the engines. Several start attempts on one engine is part of the same engine start, provided that the failure to start the engine is due to performance deviating from the checklist. If the failure to start an engine is due to the inserted error potential (hung or hot start), or because the instructor orders a new start (check the instructor score sheet), then the second attempt to start that same engine is considered a new engine start.

A new engine start is always performed (and scored) from engine checklist item no. 1. When there is a shutdown for reasons other than the inserted error potential/instructor orders (not a new start), the analyst will disregard performance in the repetition of checklist items up to the point where this start attempt failed (it does not matter if they start at no 1 or not, and if they do start at no 1, then new deviations are irrelevant.)

All deviations from the checklist must be described: A<u>ctions</u>: call outs & orders / switches & controls. I<u>tem sequences (all actions)</u> both between and within items. <u>Time (actions in bold)</u>: scored if the maximum limit is exceeded. <u>L or R pilot (P):</u> scored if their role is reversed. Type of start and possible problems are logged in the instructor pre + eng score sheets.

V = correct performance, / = missing, otherwise describe (use “other” if not enough space)

Pre Start Sample. FFS is a B type / B checklist used (AutoC OFF)

EXT. PWR	Ext.pwr ON + Voltage sel.	
BATTERIES	Battery L + R ON	
	Voltage L + R/ min 24V	
INVERTERS	MAIN Inv./ ON	
	STB Inv./ MAIN	
AVIONICS	Left/ ON	
	Middle/ ON	
	Right/ ON	
O/H	Emergency Light/ OFF	
SWITCHES	Temp. Select L/ AUTO	
	Temp. Select R/ AUTO	
	Auto Coars./ OFF	
	Boot Indicator/ ON	
	No Smoking/ ON	
LAMP/	Lamps Sw./ DOWN:cont	
SMOKE/	Lamps Sw./ UP: cont	
BLEED	Smoke Sw./ UP: cont	
	Bleed Leak/ L + R	
FIRE & FIRE	Fire Test Sw./ L + R	
SHORT	Fire Short Sw./ UP	
SHAKER &	Stall 1/ UP: cont	
PUSHER	Stall 2/ UP: cont	
	Stall 1&2/ UP: dual/cont	
	Stall 1&2/ DOWN: dual/cont	
ERROR/other		

Normal/Abnormal Engine Start Sample (always motoring start)

Checklist Items			**Description of All Deviations from Correct Performance of Checklist Items**			
Starting Left or Right1:			**LEFT ENGINE:**		**RIGHT ENGINE:**	
(start from no. 1)			START NO. 1	START NO. 2	START NO. 1	START NO. 2
1	call out "starting right"	LP				
2	call out "right prop free"	R/L				
	L/R PL ground idle L/R CL fuel off L/R ignition sw off	LP	*observe:*		*observe:*	
	if BUS TIE CONN green		*observe:*		*observe:*	
3	call out "buss tie"	LP				
	if NO BAT START out		*observe:*		*observe:*	
4	call out "NO no bat start"2					
5	hold starter sw in L/R pos.	LP				
6	**and call out "timing"**	LP				
	max 30 sec. (# 10/11)	P	*(count to 30)*		*(count to 30)*	
7	order "fuel on"	LP				
8	**L/R CL to START** and immediately	RP				
9	**call out (L/R) "fuel is on"**	RP				
10	**x ign. sw to norm ASAP3** within 2 sec. from # 8	LP				
IF 2 SECONDS EXCEEDED (Failure to Start): SHUTDOWN + MOTORING (page 8)						
11	**Release st sw (ASAP)**	LP				
12	R hand on CL	LP				

1 The trainee continues to operate the L or R system according to this callout. Callouts and actions must correspond. (if not: this is one deviation only, i.e. disregard repetition of chekl. items). Clues to detect start L/R: the handgrip on the starter switch + lights on fuel panel: L/R main pump.

2 If light comes on during engine start, continue. If light is not out before engine start, then terminate the start.

3 Otherwise: failure to start the engine.

IF ABNORMAL[4] (See Instructor Score Sheet): SHUTDOWN + MOTORING + READ CHECKL/prior to next attempt (page 8)						
13	callo. “Ng, fuel flow, ITT”	LP				
14	call out “engine oil”	RP				
15	call out “prop oil”	RP				
	lights out engine+fuel pane		*observe:*		*observe:*	
16	L/R[5]: call out “engines and fuel panels checked”	LP				
	if BUS TIE CONN green		*observe:*		*observe:*	
17	call out “bus tie”	LP				
	IF NOT GREEN:					
1	check bat.volt min 20V	LP				
18	reset R generator, 2 attemp.	LP				
	Items 19-22[6]: the <u>first</u> engine start w. <u>ext</u>. pwr					
19	ext pwr off	LP				
	white light out		*observe:*		*observe:*	
20	disconnect sign. (time out)	LP				
	ext pwr. blue light out		*observe:*		*observe:*	
	L/R gen light out		*observe:*		*observe:*	
21	load select R gen	RP				
22	call out “ load below 200”	RP				
	(- end -)					

4 ABNORMAL = SHUTDOWN (CL Fuel Off) + MOTORING (Ignition Sw. Off / Starter Sw. L/R) + READING (abnormal checklist before next start attempt).

5 Note that there is a L and R side of these panels and that only one side is checked at a time. The word “checked” means OK and it is wrong to say “checked” if any of the lights on the relevant panel side are on. If any of these lights are on, then it is correct to say noting (and to start investigating the system failure). When both system status and call out can be observed, then we score both, i.e. that the call out actually corresponds to the true system status and that the call out itself is verbally correct.

6 Item 18 is the last regular item. Item 19-20 only when used ext. pwr. Items 19-22 after starting the first engine (load is checked before start of second engine only).

	ABNORMAL:					
1	CL fuel off **ASAP**	LP				
2	ignition sw. off **ASAP**	LP				
3	**hold starter sw until safe**[7] **minimum 10 sec.**	LP				
4	Read A. Checkl. hung/hot	P				
	(wait min. 3 minutes)		(N/A)			
	Hungstart:					
	if GPU (ext pwr.)					
1	voltage ext. pwr.					
	if battery:					
1	voltage battery					
	OTHER:					

7 The starter can be engaged max 70 seconds total in one continuos sequence (each engine) and (of these 70 seconds) max 30 sec for motoring start. After one starter sequence, there must be a delay of 3 minutes. However: 1) it is not likely that motoring start requires 30 seconds, i.e. there should be more than 40 seconds left to motor the engines in case of abnormal starts and 2) if necessary: it is better to damage the starter than the turbines. MOTORING (until ITT is below 150 degrees C) or MINIMUM 10 seconds.

10.2 The Error Potential

The error potential was inserted into the checklist task (and scored in the checklist score sheet) dependent on pass number. Trainees did not know which pass that contained the inserted error potential; cockpit *configuration* during pre start/*abnormal starts* during engine start. Trainees were instructed not to discuss the error potential with other trainees.

- Pre Start: the *default configuration* was the way it was left after the last flight of the day, according to the checklist. The inserted *error potential* was (1) the deliberate erroneous pre start cockpit configuration deviating from this default (described pass by pass, see below). This created a potential for errors, specifically errors of (configuration) omission.

- Engine Start: the *error potential* was also (2) the two types of abnormal starts: (a) the *hung start* and (b) the *hot start*. Explanation: The turbo prop engines have two turbines: the gas turbine and the power turbine. The gas turbine blows hot gasses into the power turbine, which in turn drives the propeller. Hung start (a): The gas turbine is supposed to accelerate during the starter process, i.e. the DC starter generator should only drive the gas turbine up to a certain *rounds per minute* (RPM). The gas turbine RPM is measured as a percentage of propeller speed (Ng), and the gas turbine could stop around 40 percent. It may look like the engine has started, but there is no combustion and it will eventually be noticed that the engine has not started. This is called a *hung start* (also called a *hang start*) and it can be detected directly from the Ng display. Hot start (b): The temperature between these two turbines, i.e. the inter-stage turbine temperature (ITT), must not reach a value where it can damage the turbines. Different types of problems during a start, including hung start, may result in a *hot start*, which is the most serious scenario in terms of potential damage to the engine (turbines).

- Hung and hot starts look (almost) the same in the current versions of the power plant trainer and the MATE. The MATE simulation should not be different from the simulation in the power plant trainer. This had as a consequence that the attempt to measure diagnosis (engine instructor score sheets, p 2) was disregarded.

- Additional: the inserted malfunctioning of an engine, set by the instructor, constitutes a potential for errors. Another slightly complicating factor was the various types *of normal starts*, i.e. it is not an error potential, but it has some minor consequences for what actions should/should not be carried out: the engine can be started from *external power*, from *batteries* or from the engine already started (i.e. *generator*). Performance was always scored in relation to the type of start being performed, noted on the front page of the score sheet (see appendix 11.1).

Error Potential during Pre Start:

Not all the potentials for errors could be observed, i.e. the trainees' detection of malfunctioning light bulps was only written down in the instructor notes during checkout.

EXP - MATE - pre start - video log pass 1 - 4. (Instructions on how to use MATE is not included here)

- This video log gives an overview of the 4 pre start passes in MATE. The cockpit configurations to be set by the instructor are described for each pass. It can also be commented on technical / practical problems related to the video recordings.
- Before a crew starts the pre start checklist, make sure that the default configuration of switches, i.e. as the a/c was left by the previous crew (last flight of the day), is correct according to checklist and equal for all crews. During exercises it is always the first flight of the day. The only change to the default configuration is the inserted error potential.
- The inserted error potential, i.e. deviations from the default configuration, must be the same for all trainees within one pass. Make sure trainees do not know the error potential before the exercise.
- Sw positions can NOT be set from the instructor control station, but must be set manually.

PASS 1	INTRODUCTION ONLY

Default according to checklist: all mandatory sw must be off/in/guarded/safetied, EXCEPT: AutoC. and Boot indicator must be on. Emerg.L. off. Inverter main. As required sw must be off..

PASS 2			FIRST INSERTED ERROR POTENTIAL **-- COMMENTS**
Crew	Date	Ex	*Inserted error potential: 1) cargo light u/s, both bulbs (demonstration)*

PASS 3			OBSERVABLE ERROR POTENTIAL (1) **-- COMMENTS**
Crew	Date	Ex	*Inserted error potential: 1) One HP valve on, 2) No bat light u/s, both bulbs*

PASS 4			OBSERVABLE ERROR POTENTIAL (1) **-- COMMENTS**
Crew	Date	Ex	*Inserted error potential: 1) both gen sw on, 2) de-ice ovt light out, one bulb, 3) bat vent circuit breakers (N/A), 4) flight log missing (N/A)*

CHECKOUT, PASS 5:
These four potentials for errors were used during the checkout pass 5. Potential number 1-3 was for the *training sample*, potential number 4 was for the complete *checkout sample*:

1) Auto C. off
2) Win-shield ice x2 on
3) No bat. Start u/s (none of the two light bulbs working, i.e. black caution light)
--
4) Torch left side u/s, no battery, part of MEL

10.3 Error Categorisation Sheet

The following score sheet (all the tables) represent one combined pre start and engine start trial for one crew combination. Usually one crew combination performed three trials (engine starts) within one session. Ideally there should be 10 sessions within a pass (5 crews in two combinations). We looked at the first three passes with combined pre start and engine starts in the FFS.

Crew Letter LP:	Crew Letter RP:
Pass:	

Pre-start checklist FFS (config errors)			
Flow	**Omission**	**Commission**	**Total**

Engine start FFS			
Left engine:	Right engine:	Normal start:	Abnormal start:

Operational definition (v. 6.0) of Skyways extended checklist (v. 960901)	**Config errors** (switches and controls in relation to checklist and system status)	**Callout errors** (command and callout phraseology in relation to checklist and system status)	**Role errors** (reversal of whom is performing the checklist actions)
Flow			
Flow (timing)			
Omission			
Omission (transition)			
Commission			
Total			

10.4 What-To-Do-List for Instructors

These have been additional to briefing. The instructions were written in one normal (short) and one extended (detailed) version for all conditions. Camera instruction manuals were enclosed.

What-To-Do-List for Skyways FFS Instructors

SIMULTANIOUS RECORDINGS: pre st. & engine start (both running)

CHECK:

1) **Pictures** in camera monitors: overhead panel + throttles

2) **Focus**, adjust if not OK (“throttle” camera is locked)

3) **Illumination** (all light on) and exposure (AE)

COMPLETE THE SCORE SHEET FOR EACH CREW IN EACH PASS

LABEL:

1) **While recording:** state crew ID letter and pass number + show letter

2) **Write on tape:** date, instructor ID, crew ID and pass number.

10.5 Instructor Notes, pre start

EXP - A/C PRE START CHECKOUT INSTRUCTOR NOTES. Pass number 5, 6 or 7: __________ (number)

Crew ID	Date	Time	Error 1-4	Other Errors Observed?	Tot. Err.	Res. OK	COMMENTS (why OK/not OK)
/K	/	Tot.					
/L	/	Tot.					
/M	/	Tot.					
/N	/	Tot.					
/O	/	Tot.					
/P	/	Tot.					
/Q	/	Tot.					
/R	/	Tot.					
/S	/	Tot.					
/T	/	Tot.					

Maximum allowed time is 15 minutes, (i.e. a formal checkout / CTR). If trainees <u>delay</u> at checklist items that are <u>not</u> a part of MATE (only “simulated” in MATE), they must be told to continue the checkout: The instructor must monitor this during the checkout.

<u>Inserted error potential:</u>
1) Auto C. off
2) Win-shield ice x2 on
3) No bat. Start u/s (none of the two light bulbs working, i.e. black caution light)
--
4) Torch left side u/s, no battery, part of MEL; N/A for comparison EXP-CTR pass 5, but please insert in pass 7 if pass 5 & 6 fails

10.6 Instructor Score Sheets, pre start

This score sheet was produced in several versions dependent on the condition

CONFIDENTIAL - TO BE HANDED OVER TO RISØ EXPERIMENTER

Instructor score sheet for EXP (MATE) pre start CHECKOUT in a/c

Instructor initials:	
Pass no. 5, 6 or 7:	

Crew member letter K-T:	
Date:	

Please circle the number you think indicates the pilot's level of ability in each category at this stage of training. By definition, an average pilot will receive a "3".

1. Pilot's knowledge of the cockpit-layout
How well does the pilot know the location of the controls and indicators?

1	2	3	4	5	**Comments:**
low				high	

2. Speed in carrying out the pre start checklist
How fast does the pilot carry out the task (compared to other trainees)?

1	2	3	4	5	**Comments:**
low				high	

3. Task accuracy
Was there a high number of errors in carrying out the checklist? (A failure to visually inspect a parameter will also count as an error: if possible, the trainee should be asked to point at the parameter or report verbally when performing the check.)

1	2	3	4	5	**Comments:**
low				high	

4. Understanding of a/c systems: actions outside normal checklist
Knowledge of how the a/c system variables underlying the procedure are related to each other.

1	2	3	4	5	**Comments:**
low				high	

5. Predicted level of performance in the three first FFS passes (after checkout)
An estimate of how the FFS instructor will evaluate the pilot.

Lower Third	Medium Third	Upper Third	**Comments:**

Insert malfunctions/error potential before each checkout. Register on separate notes:
1) Auto C. off, 2) Win-shield ice x2 on, 3) No bat. Start u/s (none of the two light bulbs working, i.e. black caution light) 4) torch left side u/s: no battery, part of MEL; N/A for comparison EXP-CTR pass 5, but please insert in pass 7 if pass 5 & 6 fails.

10.7 Instructor Score Sheets, engine start

This score sheet was in several versions depending on the condition. Page 2 was taken out because the measure of diagnosis had to be disregarded.

CONFIDENTIAL - TO BE HANDED OVER TO RISØ EXPERIMENTER

Instructor score sheet for pre start/engine start FFS passes

Instructor initials:			Crew letters (K-T):	/
Pass no. (“Lesson SIM”) 1-7:			Date 1998:	/

Please complete this score sheet (2 pages) for each CREW in each pass.
Write N/A when an item is Not Applicable.

The trainees must not know if the start will be normal or abnormal.

Type of Normal Engine Start:

Please insert the same type of error potential, i.e. equal difficulty for all crews, within one pass.

Inserted Error Potential
Pre Start: overhead switch positions deviating from the default:
Abnormal Engine Start: type of inserted malfunction; “hung start” or “hot start”:

Please circle the number you think indicates the pilot’s level of performance.
By definition, an average pilot will receive a “3”:

1. Please indicate the level of performance on the pre-start checklist:

1 Low	2	3 Average	4	5 High

2. Please indicate the level of performance on normal engine start:

1 Low	2	3 Average	4	5 High

3. Please indicate the level of performance on abnormal engine start (hung or hot):

1 Low	2	3 Average	4	5 High

10.8 Trainees' and Instructors' Opinioins Questionnaires

Trainees' Opinions of MATE

Your time in completing this questionnaire is greatly appreciated.
Please answer within the 15th of August and deliver your answers to xxx xxx. All answers will be treated anonymously.

Crew ID letter: ________ Date: __________

We would prefer to have your answers in English, but answer in
Swedish if this is important for the accuracy of your descriptions. Use both sides of the pages if you need more space. Please make your handwriting as clear as possible.

---

1. What are your main opinions of the MATE as a training environment for the course you have just completed? Please describe the most important positive and/or negative aspects you can think of (A-C):

A) How is the MATE compared with the Paper Tiger?

B) How is the MATE compared with the Power Plant Trainer?

C) How is the MATE compared with the Aircraft (pre start)?

2. Did you consider it a problem to operate the touch screen switches? (Circle)

Yes	**No**

If yes, please describe the problem in relation to the specific switches:

3. Do you think that the MATE aspects you have described in questions 1-2 have influenced your performance in the real environments A/C and FFS? (Circle)

Yes	**No**

If yes, please describe this influence for each relevant aspect of the MATE:

4. Any other comments:

Instructors' Opinions of MATE

Your time in completing this questionnaire is greatly appreciated.
Please answer within the 15th of August and deliver your answers to xxx xxx.
All answers will be treated anonymously.

Name: ____ Date: __________

Please circle your direct and/or indirect experience with MATE as an instructor:

A/C checkout - MATE - FFS - Other, specify:____________________

If you do not have any MATE relevant instructor experience (directly in MATE or indirectly in the A/C checkout or FFS), but you do have some knowledge of the MATE (Other, specify), then please comment on a general basis. Write N/A (Not Applicable) if you do not have any opinions of the MATE. We would prefer to have your answers in English, but answer in Swedish if this is important for the accuracy of your descriptions. Use both sides of the pages if you need more space. Please make your handwriting as clear as possible.

---

1. What are your main opinions of the MATE as a training environment for the Saab 340 course at Skyways? Please describe the most important positive and/or negative aspects you can think of (A-C):

A) How is the MATE compared with the Paper Tiger?

B) How is the MATE compared with the Power Plant Trainer?

C) How is the MATE compared with the Aircraft (pre start)?

2. Do you consider it a problem to operate the touch screen switches? (Circle)

Yes **No**

If yes, please describe the problem in relation to the specific switches:

3. Do you think that the MATE aspects you have described in questions 1–2 influenced the trainees' performances in the real environments A/C and FFS? (Circle)

Yes **No**

If yes, please describe this influence for each relevant aspect of the MATE:

4. How do you think that MATE can improve pilot education and/or training?

Please take cost-effectiveness into consideration and also include comments and suggestions related to:

- Education/training of pilots that are <u>new</u> to a specific a/c type.
- Education/training of pilots with <u>experience</u> on a specific a/c type (e.g. CRM).
- Any suggestions for the further development and use of MATE.

5. Any other comments:

Bibliographic Data Sheet **Risø-R-1324(EN)**

Title and authors

MATE, Multi Aircraft Training Environment

G. Hauland, T. Bove, H.B. Andersen, S. Weber (Risø)
R. Eriksson and G. Fletcher (DERA)

ISBN	ISSN
87-550-3015-7 (Internet)	0106-2840

Department or group	Date
Systems Analysis Department Man/Machine Interaction	January 2002

Groups own reg. number(s)	Project/contract No(s)
ESPRIT PROJECT 8982	

Pages	Tables	Illustrations	References
68	3	16	3

Abstract (max. 2000 characters)

A medium fidelity and low cost training device for pilots, called the Multi Aircraft Training Environment (MATE), is developed to replace other low fidelity *stand-alone* training devices and integrate them into a flexible environment, primarily aimed at training pilots in checklist procedures. The cockpit switches and instruments in MATE are computer-generated graphics. The graphics are back projected onto semi-transparent touch screen panels in a hybrid cockpit mock-up. Thus, the MATE is relatively cheap, it is always available, it is reconfigurable (e.g. between types of aircraft/models to be simulated) and with possibilities for including various forms of intelligent computer assistance. This training concept and the technology are not specific to aviation, but can be used to simulate various types of control panels in different domains. The training effectiveness of pilots' procedure training in the MATE prototype was compared with the effects of traditional training that included the use of real aircraft. The experimental group (EXP) trained the pre-start checklist and the engine start checklist for the Saab 340 commuter aircraft in a MATE prototype. The control group (CTR) trained the same procedures using the aircraft (a/c) for training the pre start and a desktop computer tool (power plant trainer) for training engine starts. Performance on the pre-start checklist was compared in a formal checkout that took place in the a/c. Performance on the engine start procedure was compared in a full flight simulator (FFS). The conclusion was, firstly, that training in the MATE prototype can result in an equally good performance as the existing training (a/c and computer tools), provided that the MATE trainees are given time to familiarise themselves with the a/c. Secondly, training in MATE can result in better performance during dynamic tasks, such as abnormal engine starts. This is promising for the further development of the MATE concept.

Descriptors INIS/EDB